Vol. 49 FEBRUARY, 1958 No. 1

The Journal of Educational Psychology

CONTENTS

Patterns of Personal Problems of Adolescent Girls 1
RICHARD B. SCHUTZ

Readability Level and Differential Test Performance: A Language Revision of the Study of Values 6
JEROME LEVY

The Attitudes Students Assign to Their Teacher 13
NORMAN M. CHANEKY

The School Progress and Adjustment of Underage and Overage Students 17
CLYDE J. BAER

A Further Note on Basal Metabolism and Academic Performance 20
MARY E. YARBROUGH AND HAROLD G. McCURDY

An Experimental Prognostic Test for Remedial Readers 23
RICHARD W. WOODCOCK

Effect of an Audio-Visual Phonics Aid in the Intermediate Grades 28
CAROLYN LUSKIN, EILEEN STANTON, AND CHARLES L. DOYLE

Predicting School Achievement for Bilingual Pupils 31
JAMES G. COOPER

The Effect of Child Psychology on Attitudes Toward Parent-Child Relationships 37
FRANK COSTIN

The Differential Effect of Manifest Anxiety on Test Performance 43
EVAN W. PICKREL

Some Socioeconomic Correlates of Academic Aptitude 47
ZEROBROW MOHANDESSI AND PHILIP J. RUNKEL

AMERICAN PSYCHOLOGICAL ASSOCIATION

The Journal of Educational Psychology

RAYMOND G. KUHLEN, Editor
Syracuse University

Advisory Editors

LEE J. CRONBACH
University of Illinois
CHESTER W. HARRIS
University of Wisconsin
ERNEST R. HILGARD
Stanford University
NICHOLAS HOBBS
George Peabody College for Teachers
HAROLD E. JONES
University of California
IRVING LORGE
Teachers College, Columbia University
DOROTHEA McCARTHY
Fordham University
T. ERNEST NEWLAND
University of Illinois
H. H. REMMERS
Purdue University
JULIAN C. STANLEY
University of Wisconsin
J. B. STROUD
State University of Iowa
DONALD E. SUPER
Teachers College, Columbia University
PERCIVAL M. SYMONDS
Teachers College, Columbia University
GEORGE G. THOMPSON
Syracuse University

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Circulation Manager
SARAH WOMACK, Editorial Assistant

The Journal of Educational Psychology publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association*. Because of the great demand for journal space, authors should give special attention to reducing the length of manuscripts while maintaining clarity of expression. Ordinarily, papers should not exceed 15 typed pages. Accepted manuscripts are published in order of their receipt. An author may, however, request early publication of accepted manuscripts if he is willing to assume the cost of printing the necessary extra pages.

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D. C. Address changes must reach the APA by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $8.00 (Foreign $8.50). Single copies, $1.50.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1333 Sixteenth Street N.W., Washington 6, D. C.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the act of March 3, 1879.

Copyright, 1960 by the American Psychological Association, Inc.

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 49 | February, 1958 | Number 1

EDITORIAL NOTE

This is the first issue of the first complete volume of the *Journal of Educational Psychology* to be published by the American Psychological Association and the first issue under a new editor. Delays and other matters incident to the transition of the journal to *both* a new publisher and a new editor have resulted in certain temporary departures from usual policy and some delay in the publication of first issues. Contributors and readers will note in this and the next several issues, for example, certain inconsistencies in style and certain departures from standard practice of publishing manuscripts in order of their receipt. In the future, contributors should follow the style prescribed by the *Publication Manual of the American Psychological Association.*

R. G. K.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 1, 1958

PATTERNS OF PERSONAL PROBLEMS OF ADOLESCENT GIRLS[1]

RICHARD E. SCHUTZ[2]
Teachers College, Columbia University

Numerous investigators have determined the frequency with which various problems are listed by given samples of adolescents (**4**). However, in tabulating the problems it has been necessary to rely on a priori systems of classification. The customary procedure has been to categorize the problems by activity or functional areas; e.g., school, home, health, etc.

The purpose of the present study was to determine the pattern or structure underlying the personal problems which adolescents recognize and are willing to report on a youth problems inventory. This pattern was investigated by extracting homogeneous clusters from a sample of 156 items selected from the inventory.

PROCEDURE

The inventory used in the study was the Billett-Starr Youth Problems Inventory, Senior Level (**2**), a check list intended to provide a means of systematically identifying the personal problems of individual adolescents. The 441 items which make up the Inventory include problems mentioned in the compositions and free responses of several large samples of high school students which the authors obtained in developing the instrument. The items are organized into 11 areas designated as follows:

1. Physical Health, Fitness, and Safety
2. Getting Along with Others
3. Boy-Girl Relationships
4. Home and Family Life
5. Personal Finance
6. Interests and Activities
7. School Life
8. Heredity
9. Planning for the Future
10. Mental-Emotional Health and Fitness
11. Morality and Religion

The cluster analysis was based on the responses of 500 girls in Grades 10 and 11 in two Pinellas County, Florida, high schools who took the Inventory as part of the national standardization program in May 1956. The schools are three-year high schools and had a segregated white enrollment at the time of the study. The Inventory was administered in regular classrooms by regular teachers. The Inventories were signed by the students.

The basic technique of analysis was that described by Loevinger, Gleser, and Dubois (**6**) for deriving clusters which have maximum reliability as estimated by Kuder-Richardson Formula 20. Each cluster is obtained by starting with a triad of items having the highest covariance and adding items in succession, adding always the item for which the ratio of the sum of covariance with the items already in the cluster is a maximum. Items are added to the cluster until no more items remain which will increase this ratio. The process is repeated on the residual pool of items to form the second and subsequent clusters.

[1] Based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. at Columbia University. The investigation was carried out under the helpful direction of Robert L. Thorndike. The writer wishes to thank Roger T. Lennon and the staff of the Division of Test Research and Service, World Book Co. for making the study possible.

[2] Now at Arizona State College, Tempe, Arizona.

A sample of 156 items was selected to be included in the cluster analysis (**8**, pp. 57–64). The items selected have the following characteristics: (*a*) they are included in both the Junior and Senior levels of the Inventory; (*b*) they were rated as "very serious" or "moderately serious" problems by a panel of 20 guidance specialists; (*c*) each was marked by at least 5% of the *Ss* in the sample. An attempt was made to make the sample of items representative of the 11 areas of the Inventory and to include as many items as possible which have counterparts in other published problems check lists.

The Inventory attempts to get at the intensity of a student's problems by allowing him to differentiate between those which bother him "some" and those which bother him "very much." For the present analysis each *S*'s "some" and "much" responses were combined into a single category.

Each *S*'s responses were multiple punched on an IBM card, and the 156 by 156 co-occurrence matrix was prepared using the counting sorter. The figures in the co-occurrence matrix were converted to percentages and the variance-covariance matrix prepared. The cluster analysis was performed, and a check was made on the factorial purity and reliability of the obtained clusters.

Results

Three clusters were extracted from the pool of 156 items. Eighty-three items are included in Cluster I, 16 in Cluster II, and 17 in Cluster III. An abbreviated Cluster I, consisting of 37 items, was formed by eliminating 35 items which correlated less than .40 with the complete cluster and 11 items which nearly duplicated another item in the cluster; e.g., the two items, "I'm often restless," and "I'm restless most of the time." The items in the abbreviated Cluster I and in Clusters II and III are shown in Tables 1–3. The items are arranged in order of the magnitude of the point biserial correlation of each item with its cluster. The area and item number within the area are indicated in the first column of each table. The Kuder-Richardson Formula 20 reliabilities of the clusters and their intercorrelations are shown in Table 4.

The nature of a cluster must be determined by examining the items to discover the general attribute they seem to hold in common. The items in Cluster I cover a broad area, coming from eight areas of the Inventory. The cluster appears to reflect a general feeling of personal anxiety and insecurity.

The items in Cluster II are currently classified under seven different area headings in the Inventory. They seem to involve a feeling of nervous tension concerning relationships with other persons. This cluster is the least homogeneous of the three, and its correlation with Cluster I is nearly as high as its reliability. Both Clusters I and II reflect personal anxiety. While the items in Cluster II did not have enough in common with the items in Cluster I to be included in the more homogeneous general cluster, they shared sufficient common variance to form another with lower reliability.

Cluster III is the only cluster that does not cut across the functional area organization of the Inventory to any great extent. Fifteen of the 17 items in the cluster come from Area IV of the Inventory headed "Home and Family Life." The items all represent some kind of difficulty in getting along with parents.

If a cluster is factorially pure, all of its common factor variance should be accounted for by a single centroid factor. The ratio of the first factor variance to the common factor variance thus provides a basis for evaluating the factorial purity of a cluster.

A complete centroid analysis (**9**) was performed independently on the items in the abbreviated Cluster I and in Clusters II and III. The highest correlation coeffi-

TABLE 1

CLUSTER I

Item No.	r	Item
10–38	.59	People don't understand me.
10–10	.57	I'm afraid of making mistakes.
10–15	.55	I'm often restless.
10–43	.51	I worry about what others say.
10–37	.51	I feel I'm not wanted.
10–17	.50	I'm disgusted with myself (dislike myself very much).
10–19	.49	I need someone to give me advice.
8–3	.49	I would like to be able to do something well.
8–1	.49	I don't understand myself.
11–10	.49	Many times I don't know what is right and what is wrong.
10–39	.48	People talk about me behind my back.
10–2	.48	I feel uncertain (unsure) about everything.
10–34	.47	I spend too much time daydreaming.
10–3	.46	I need to learn to depend on myself.
10–35	.45	I feel sorry for myself.
10–46	.45	I get excited too easily.
11–6	.45	I'm sometimes troubled by immoral (bad) thoughts.
7–5	.45	I wonder if I'll pass.
10–32	.44	I don't get out and go after what I want.
9–18	.44	I wonder if I'm taking the right subjects.
7–44	.44	I'm afraid to take tests.
10–58	.44	I'd like to know how to get rid of a bad habit.
4–46	.44	I'm unhappy at home.
7–62	.43	Some teachers never encourage or help me.
6–1	.43	I seldom have anything interesting to do.
7–85	.43	I would like to know how to get along with certain teachers.
8–2	.43	I wonder what my real mental ability is.
10–13	.42	I wonder what my future will be.
10–1	.42	I'm confused by the way things change.
2–50	.42	I feel lonely most of the time.
4–40	.42	I'm afraid to tell my (father) (mother) when I've done something wrong.
10–42	.42	I'm blamed for things that aren't my fault.
2–52	.41	I find it hard to make friends.
10–30	.41	I don't know how to (pay attention) (work or study hard).
11–5	.41	I often tell lies.
10–45	.40	I'm bothered by people who find fault with me.
10–47	.40	I can't control my temper.

cient in each column of the matrix was used as the communality estimate, communalities being re-estimated by this method for every residual matrix. Factoring was considered complete when both Tucker's Phi and Coombs' criterion (3) indicated that the last significant factor had been extracted. Three factors were extracted from Cluster I, three from Cluster II, and two from Cluster III. First factors account for 80%, 69%, and 85%, respectively, of the common factor variance of the clusters.

To investigate the extent to which the reliability of the clusters is dependent on chance factors, the Kuder-Richardson Formula 20 reliability of each cluster was computed, based on the results of a new sample of 73 *Ss* selected from the same population as the original sample. The reliability coefficients for the new sample were as follows: Cluster I, .94; abbrevi-

TABLE 2

CLUSTER II

Item No.	r	Item
2–11	.63	I'm nervous when I talk to people.
2–15	.60	I'm not good at talking with people.
7–38	.55	I'm nervous in front of the class.
1–19	.49	I get tired easily.
2–10	.48	(I'm afraid) (I don't like) to meet people.
2–39	.47	I want others to like me.
1–21	.44	I'm always nervous.
3–1	.44	I don't understand (boys) (girls).
2–6	.38	I'm not good-looking.
6–10	.36	I would rather be alone.
6–9	.36	I get tired from too much activity.
6–11	.36	I spend too much time on (radio) (television) (movies).
1–22	.32	I need to know more about sex (body changes at my age) (new body functions).
10–40	.31	I'm afraid I seem conceited (stuck-up).
9–25	.30	I'm not sure whether I should go to college.
2–7	.28	I (don't have) (don't know how to pick) the right clothes.

TABLE 3

CLUSTER III

Item No.	r	Item
4–31	.64	My (father) (mother) is always criticizing (blaming) (nagging) me.
4–43	.59	I can't discuss things with my (father) (mother).
4–33	.58	My (father) (mother) is always expecting too much of me.
4–37	.54	I'm often the cause of family quarrels (parents argue about things I do).
4–51	.52	I'm thinking of leaving home.
4–41	.52	My (father) (mother) has little or no interest in what I do.
4–39	.51	I sometimes lie to my (father) (mother) to get permission to do something.
4–60	.51	My (father) (mother) is often nervous and irritable.
4–42	.50	My (father) (mother) never asks my opinion about anything important to the family.
4–50	.49	I dislike my (father) (mother) very much.
4–25	.49	I don't agree with my (father) (mother) about my out-of-school activities.
4–48	.48	I'm sometimes ashamed of things my parents do or say.
4–38	.45	My (father) (mother) pries into my private affairs.
4–26	.44	My (brother) (sister) is always causing me trouble.
5–4	.43	I don't get an allowance.
2–19	.40	I often "stretch the truth" when I tell something.
4–49	.34	There's too much drinking in our home.

ated Cluster I, .91; Cluster II, .63; Cluster III, .83.

The significance of the difference between each of the coefficients for the new and the original group was computed using Fisher's z transformation. The difference for Cluster II is significant at the .001 level. There is no significant difference

TABLE 4
CLUSTER INTERCORRELATIONS AND KUDER-RICHARDSON RELIABILITIES

Cluster	I(83)	I(37)	II	III
I (83)	(.94)	.98	.67	.60
I (37)	.98	(.90)	.63	.52
II	.67	.63	(.71)	.34
III	.60	.52	.34	(.81)

between the coefficients for the other clusters.

DISCUSSION OF RESULTS

The findings of the study have implications for both the theory and measurement of adolescent problems. While the study was not designed to test the validity of the various theories of adolescent problems that have been proposed (**1, 5, 7**), it does provide evidence concerning the way in which problems cluster or "go together."

The cluster structure does not correspond closely to any of the theoretical frameworks which have been proposed for classifying adolescent problems. The composition of Cluster I suggests that a single dimension of personal anxiety underlies many of the manifest problems which theorists have used several dimensions to explain.

The fact that two of the clusters cut across several areas of the Inventory indicates that the classifying of items in problems check lists into the traditional functional or activity categories is in large part an arbitrary procedure. However, even though the conventional rubrics do not all represent true functional unities, teachers and counselors may still find the categories helpful. The subdivisions often suggest programs of action related to the kinds of services that schools and other social agencies are equipped to provide. Thus, the traditional area organization often serves the purpose of suggesting foci of therapeutic action.

Although no data are yet available concerning the predictive validity of the clusters, they provide a potential means of screening adolescents in need of psychological help. They have high face validity and are reasonably pure factorially. A study of predictive validity is required to determine the extent to which they are actually related to personal adjustment.

SUMMARY

A cluster analysis was performed on 156 selected items from the Billett-Starr Youth Problems Inventory, Senior Level, based on the responses of 500 adolescent girls. Three clusters were extracted and designated as follows: Cluster I—General personal anxiety and insecurity; Cluster II—Tension concerning relations with others; and Cluster III—Difficulties in getting along with parents. The implications of the cluster structure for the theory and measurement of adolescent problems was briefly discussed.

REFERENCES

1. AUSUBEL, D. P. Problems of adolescent adjustment. *Nat. Assoc. sec. Sch. Prin. Bull.*, 1950, **34** (167), 1–84.
2. BILLETT, R. O. & STARR, I. S. *Billett-Starr Youth Problems Inventory.* Yonkers-on-Hudson, N. Y.: World Book, 1956.
3. FRUCHTER, B. *Introduction to factor analysis.* New York: Van Nostrand, 1954.
4. GRANT, B. Survey of studies on problems of adolescents. *Calif. J. sec. Educ.*, 1953, **28,** 293–297.
5. HAVIGHURST, R. J. *Human development and education.* New York: Longmans, Green, 1953.
6. LOEVINGER, JANE, GLESER, GOLDINE C. & DUBOIS, P. H. Maximizing the discriminating power of a multiple-score test. *Psychometrika,* 1953, **18,** 309–317.
7. LUCHINS, A. S. On the theories and problems of adolescence. *J. genet. Psychol.*, 1954, **85,** 47–63.
8. SCHUTZ, R. E. Patterns of personal problems of adolescent girls. Unpublished doctoral dissertation, Columbia Univer., 1957.
9. THURSTONE, L. L. *Multiple factor analysis.* Chicago: Univer. of Chicago Press, 1947.

Received August 10, 1957.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 1, 1958

READABILITY LEVEL AND DIFFERENTIAL TEST PERFORMANCE: A LANGUAGE REVISION OF THE STUDY OF VALUES

JEROME LEVY[1]
Denver University

Psychologists have recently become interested in the "readability" of material; that is, how the verbal difficulty of written material affects the communication of ideas. Hebb and Bindra (**5**), Ogdon (**7**), and Stevens and Stone (**8**) have published material relevant to this question. These writers have been concerned with readability in discursive writings; e.g., textbooks. The important question, however, of how the readability level of a psychological test might affect the score which a subject with a defined level of verbal competence achieves has not hitherto been directly investigated. It appears to the writer that this is a question of considerable importance to sound psychometric practice. The present study was designed to investigate this question.

The Allport-Vernon Study of Values, 1951 Revision (**1**), was selected as an appropriate instrument. Since its introduction in 1931 and following its revision in 1951, the Study of Values has been employed in researches in personality theory, and for clinical evaluation and guidance purposes, and has proved quite useful in these areas. Many writers have noted, however, that the level of language employed in the Study of Values is quite complex and difficult. Gough (**4**), for example, writes: "The language used is too academic and involved for use in groups very far removed from a scholastic environment." Allport notes: "The scale is designed primarily for use with college students or with adults who have had college (or equivalent) education" (**1**). Although it is not explicitly stated, one surmises that the basis of this view is the recognition of the verbal complexity of the test, particularly in terms of vocabulary usage. If, however, verbal difficulty is the limiting factor as seems suggested, then the Study of Values may not be a valid test even for *all* college students. Classes in remedial reading, basic communication, and other related areas in every university attest to the large numbers of students deficient in these verbal skills.

The present study represents an attempt to produce a modification of the Study of Values congruent in all respects with the 1951 Revision, except that it employs less difficult language. Such a modification may be useful in two ways: if it can be demonstrated equivalent to the 1951 Revision it may be validly used with *Ss* for whom the 1951 Revision is valid, and may also permit extension of the test to populations hitherto considered inappropriate (i.e., noncollege level *Ss*). Secondly, it may be utilized in demonstrating the effects of differential vocabulary ability on test scores.

PROBLEM

There are four major phases to this research: (*a*) actual construction of the modification; (*b*) a test of the meaning equivalence of the items of the modification and the 1951 Revision; (*c*) a demonstration that the language level of the modification is indeed lower than that of the 1951 Revision; and (*e*) a demonstration that there will be differences in performance between the two forms for *Ss*

[1] Now Chief Psychologist, Division of Mental Health, Colorado State Department of Public Health.

whose vocabulary ability is below that inherent in the 1951 Revision which will be *significantly greater* than the between-form differences for *Ss* whose vocabulary ability is equal to that of the 1951 Revision. This is the major focus of the investigation and the primary hypothesis to be tested. Consequent hypotheses for testing are: (*a*) For a high vocabulary ability group the between-form performance differences will be of a minimal nonsignificant nature; (*b*) For a low vocabulary ability group there will be a large and statistically significant difference in performance on the two forms; and (*c*) For a group of *Ss* intermediate on the vocabulary criterion, the between-form differences will fall between that for the high and that for the low groups.

Method

Construction of the modification. Of the 45 items in the 1951 Revision, all but one were reworded. Items were constructed which attempted to maintain the structure and meaning of the equivalent original items, but which were worded much more simply. Where a choice of words with equivalent meanings was to be made, the simpler word was always chosen.

Equivalence of meaning. Both the 1951 Revision and the modified form were given to a group of judges with a rather heterogeneous background in psychology. These judges were unfamiliar with the Study of Values. They were supplied with the definitions of the six predominant personality types as they appear in the Manual of Directions (**1**) so that they would have clear referents for each scale. They were presented the two forms in mixed order and asked to judge which type of person would answer each question with each alternative choice. For example, in the item "Assuming you have sufficient ability, would you prefer to be: (a) a banker; (b) a politician?" which of Allport's types would choose alternative (a), which would choose (b)? The judges also matched the items of both forms by number; the items of each form having been randomized.

Readability. In order to evaluate the hypothesis that the language level of the modified form was simpler, a test of "readability" was employed. This is the Flesch formula (**3**).

Between-form differences. A group of 157 young male Air Force *Ss* was administered the 1951 Revision, the present modification, and the Diagnostic Reading Test, Survey Section (**2**). The two forms of the Study of Values were given in balanced order of administration, with the Diagnostic Reading Test coming between them for all *Ss*. The vocabulary scale of the Survey Section was taken as the primary criteria for the selection of three groups: a *Low* group of 22 *Ss* whose vocabulary scores placed them at or below the lower quartile of the eighth grade, a *High* group of 23 *Ss* with vocabulary scores at or above the upper quartile of the twelfth grade, and a *Middle* group of 20 *Ss* with intermediate vocabulary scores. Thus there resulted three groups of approximately equal size, with no overlap on the vocabulary criterion, whose tested vocabulary ability closely corresponded to the readability of the two forms of the Study of Values. The two forms were scored for each *S*, and a "deviation score" was also computed. This is merely the difference between *S*'s score on a scale on one form and his score on the same scale on the alternate form, either positive or negative (sign being dropped), and summed for all six scales.

Results

The results of the judgments indicate that, in the opinion of these judges, the revised items *do* ask the same thing as the original items. The judges made the same categorizations of the original and the revised items; i.e., there was not a statistically significant difference between

TABLE 1
SUMMARY OF *t* TESTS OF ORDER FOR THE HIGH, MIDDLE AND LOW GROUPS

Theoretical	Economic	Aesthetic	Social	Political	Religious
Low Group					
t[a] .11	1.12	.65	.08	.29	.07
Middle Group					
t[b] .98	1.07	.11	.44	.80	1.42
High Group					
t[c] .64	.26	.52	.58	.74	.59

[a] .05 = 2.086, *df* = 20
[b] .05 = 2.101, *df* = 18
[c] .05 = 2.080, *df* = 21

the type judgments on the pairs of items. The results of the item-to-item match show that, of the total of 506 matches the judges made, 495—or 98%—were matched correctly.

The Flesch formula yielded a "Reading Ease Score" of 52.7 for the 1951 Revision, placing it at the twelfth grade level: the score of 72.6 for the modified form is at the seventh grade level.

TABLE 2
ANALYSIS OF VARIANCE TO DETERMINE SIGNIFICANCE OF BETWEEN-FORM DIFFERENCES FOR THE LOW GROUP, AS COMPARED TO BETWEEN-FORM DIFFERENCES FOR THE HIGH GROUP

	Sum of Squares	*df*	Mean Square	*F*
Between	609.158	2	304.579	3.955
Within	4,774.780	62	77.103	
Total	5,383.938	64		

	Low Group	Middle Group	High Group
Mean Differences	24.182	22.150	17.043

Since some of the *Ss* in each of the three groups took the forms in original-modified order, and others took them in modified-original order, *t* tests were computed to determine whether the order in which the forms were taken has a significant effect on the between-form differences. As may be seen from Table 1, none of these *t*'s was significant at the .05 criterion level. It was therefore possible to combine all of the *Ss* within each of the three groups, and to treat each of the groups as a single unit for further statistical evaluation.

As a test of the main hypothesis—that is, that the Low group would show significantly greater between-form differences than would the High group—an analysis of variance was performed utilizing the deviation scores. The *F* value of 3.955 which was obtained, as shown in Table 2, is significant at, and beyond, the .05 level of confidence. This clearly substantiates the major hypothesis. The primary characteristic on which these two groups differ is that of diagnosed vocabulary ability, and the significantly greater between-form differences in the

Low group seems clearly related to their low vocabulary level. The group means of the difference scores, which are also contained in Table 2, bear out the hypothesis that as the more difficult form moves away from the *Ss*' level of vocabulary ability they make increasingly different scores than they do on the form which is within their level of competence.

The groups show a good deal of consistency in mean scale values from one form to another, as indicated by the data in Table 3. On the Theoretical scale, the Low group scores lowest on both forms, the Middle group is intermediate on both forms, and the High group scores highest on both forms. The Aesthetic scale shows the reverse order, with the same relative consistency on both forms. The *t* tests were computed to determine the significance of change in performance on each of the six scales for the High and Low groups: these data are presented in Table 4.

The hypotheses employed in this investigation state, in effect, that the two forms of the Study of Values are parallel

TABLE 3

MEAN SCALE VALUES ACHIEVED BY THE LOW, MIDDLE AND HIGH GROUPS ON THE 1951 REVISION AND THE MODIFICATION

Group	Theoretical	Economic	Aesthetic	Social	Political	Religious
Low						
1951 Revision						
Mean	43.136	40.272	34.181	40.454	42.772	39.181
Modified Mean	40.136	42.363	33.954	41.045	42.454	40.045
Middle						
1951 Revision						
Mean	44.500	43.550	32.350	35.250	43.350	41.000
Modified Mean	44.950	42.150	33.600	36.850	41.300	41.150
High						
1951 Revision						
Mean	49.347	44.826	30.913	33.173	40.260	41.378
Modified Mean	47.739	44.304	31.478	35.478	40.260	40.739

TABLE 4

t TESTS OF SIGNIFICANCE OF DIFFERENCES IN PERFORMANCE BETWEEN 1951 REVISION AND MODIFICATION ON EACH OF THE SIX SCALES FOR HIGH AND LOW GROUPS

	Theoretical	Economic	Aesthetic	Social	Political	Religious
Low Group						
t	3.86	2.29	.37	1.36	1.13	1.77
P	.001	.05	NS	NS	NS	NS
High Group						
t	1.47	1.18	1.55	2.62	.46	1.97
P	NS	NS	NS	.02	NS	NS

TABLE 5

TEST-RETEST CORRELATION COEFFICIENTS FOR THE 1951 REVISION AND OBTAINED CORRELATION COEFFICIENTS BETWEEN PERFORMANCE ON THE 1951 REVISION AND THE MODIFIED FORM FOR THE HIGH AND LOW GROUPS

	1951 Revision (Test-Retest)	Low Group correlations between 1951 Revision and Modified	High Group correlations between 1951 Revision and Modified
Theoretical	.87	.60	.76
Economic	.92	.53	.84
Aesthetic	.90	.60	.90
Social	.77	.74	.68
Political	.90	.55	.89
Religious	.91	.72	.91

Tests of Significance Between Correlation Coefficients

	High to Low Critical Ratio	High to 1951 Revision Critical Ratio	Low to 1951 Revision Critical Ratio
Theoretical	1.27	1.35	2.68[a]
Economic	2.65[a]	1.47	4.18[c]
Aesthetic	3.27[b]	0.00	3.26[b]
Social	.51	.76	.79
Political	3.18[a]	.30	3.57[b]
Religious	2.60[a]	0.00	2.59[a]

[a] .01 level
[b] .001 level
[c] .0001 level

forms for the High vocabulary group where there are not the interfering effects of vocabulary limitations, and are not parallel forms for the Low group where these limitations do exist. If the two forms are parallel, the correlation between performance on the Modification and on the 1951 Revision should not differ significantly from a test-retest correlation of performance on the 1951 Revision alone. In other words, if the two forms are equivalent there should not be a statistically significant difference between the correlation coefficients obtained by giving the 1951 Revision twice, and those obtained by correlating performance on the two forms. Pearson product-moment correlation coefficients were computed to determine the degree of relationship of the two forms in the High and Low groups. To determine whether these *r*'s are significantly different from each other the r-to-z transformation and critical ratio test of significance was employed. These data are presented in Table 5. Because of the impossibility of obtaining test-retest data on this group of *Ss*, these data are taken from the 1951 Revision Manual (**1**).

DISCUSSION

It seems reasonable to expect that people who are highly theoretically oriented are unlikely to have a very low vocabulary level. Omitting any question of cause and effect, it does seem likely that there is a positive relationship between these areas. People who are theoretically oriented are likely to spend a good deal of time reading in many areas of knowledge, and in other ways are apt to develop a well-organized and comprehensive vocabulary. Conversely, persons not theoretically oriented may reasonably

be expected to engage less in these activities. What the writer proposes, in effect, is that the higher mean Theoretical score found on the original form is *spuriously high*. When items are not understood, choices are likely to be made on the basis of something other than true preference, and it is suggested that the higher mean for this group on the original form is due largely to noncomprehension. When, on the other hand, they are presented the same alternative choices in language that they can understand, they tend to score lower. Since as a group they are not highly theoretically oriented, the group mean score on this scale will fall as the meaning of the items become clear. If this view is correct, then the modified form is indeed a more valid measure of theoretical values in this group.

The lower Economic scale scores for the Low group on the 1951 Revision is significant at the .05 level. It is suggested that this result again reflects the effects of noncomprehension. It is hypothesized that when choices having to do with economic values are presented to *Ss* in the Low group in language that is meaningful to them, they score higher on the Economic scale because this reflects a "true" underlying orientation in this direction.

The results shown in Table 4 also indicate that there is one scale on which the High group *Ss* make significantly different scores on the two forms: the Social scale. Two tentative explanations may be advanced for the significantly higher score on the modified than on the original form: (*a*) this may be a chance phenomenon, or (*b*) real differences may arise on this scale in a highly verbal group as a result of change in wording. A cross-validational study to investigate the meaning of this result is necessary.

Interpretation of the results of this study seems simple and straightforward. Since none of the between form correlations for the High group differs significantly from the test-retest correlation coefficients, the modified form is considered a parallel form for this group. For those *Ss* where there are not the vocabulary limitations, the modification yields results that do not differ significantly from that which would be obtained through the use of the 1951 Revision, and the two forms may be considered equivalent for this group. However, the significant CRs obtained in the comparison for the Low group indicate that the modification is *not* a parallel form for this group. Since the major characteristic on which the groups differ, and the basis on which they were selected, is differential vocabulary ability, it seems clearly evident that it is the vocabulary proficiency of the *Ss* which determines differential performance on the two forms. If the *Ss* can understand the language, then the modification is a parallel form of the Study of Values. If, however, the *Ss*' language ability is not sufficient to cope with the 1951 Revision, the differences in performance on the two forms are so great as to indicate that they represent different tasks.

Since the modification is a parallel form for highly verbal *Ss*, it is likely that it is as valid for these *Ss* as is the 1951 Revision. It is the writer's opinion that the modification is a *more valid* form for low verbal people. Since the level of language is brought within the comprehension of these low verbal *Ss*, random response choices based on noncomprehension are likely to be decreased. Responses based on an understanding of what the question is really asking are more likely to reflect truly the *S*'s orientation on the dimension being measured. External validation studies with different occupational groups, with other test criteria, and with generalized measures of personality are necessary for an empirical demonstration of validity. If on further investigation this modification receives external validity support, it may well serve a useful function in psycho-

logical evaluation. It will provide the same sort of personality information as does the 1951 Revision for highly verbal *Ss* and will permit extension of the test to populations hitherto considered inappropriate.

The factor of verbal level as an important variable in pencil-and-paper psychological test results has, of course, been realized for some time. But this has been at a rather gross level. This is the first study, to this writer's knowledge, which has attempted to manipulate verbal facility as an experimental variable and to evaluate statistically the effect which this has on test performance. With the growing concern about readability and communication in general, it is not likely to be the last. A good deal more study remains to be done with the modification. An item analysis is planned for the future. A covariance analysis of the relationship between differential test performance, verbal level, and intelligence is likely to clarify other important variables. But a promising start has been made in producing a meaningfully equivalent, yet less verbally complex, form of this popular test, and it has been possible to demonstrate the significant role which verbal facility plays in pencil-and-paper test results.

Summary

This paper reports an attempt to produce a meaningfully equivalent, but less verbally complex, form of the Study of Values, and to demonstrate the significant role which language facility plays in determining score patterns on pencil-and-paper test of personality. Pilot data with the modified form indicates that it is judged as asking the same things as the Study of Values. Flesch counts of the 1951 Revision and of this modification indicate that they are at the twelfth grade level and seventh grade level, respectively. Three experimental groups were established: a Low vocabulary group, a High vocabulary group, and a Middle group. The two forms were administered to these *Ss*. Analysis of the data indicates that for the High group the modification is an equivalent form. The *Ss* in the Low group make significantly different scores on the two forms, and the forms are not equivalent for this group. The differences in performance in this group are attributed to a vocabulary level inadequate to deal with the 1951 Revision, and it is suggested that for these *Ss* the modification provides a more valid test of value orientations. Much additional exploratory work remains to be done with the modification before it may be completely accepted as a valid, equivalent form for all *Ss*. This study serves to emphasize the important role of language facility in all pencil-and-paper personality tests.

REFERENCES

1. Allport, G. W., Vernon, P. E., & Lindsey, G. *Study of Values (Rev. ed.): Manual of Directions*. Boston: Houghton Mifflin, 1951.
2. Committee on Diagnostic Reading Tests, Inc. Description of the purposes and functions of the Diagnostic Reading Tests. *Educ. psychol. Measmt.* 1948, **8**, 3–14.
3. Flesch, R. *How to test readability*. New York: Harper, 1951.
4. Gough, H. G. In O. K. Buros (Ed.), *The fourth mental measurements yearbook*. Highland Park, New Jersey: Gryphon Press, 1953. Pp. 156–157.
5. Hebb, D. C., & Bindra, D. Scientific writing and the general problem of communication. *Amer. Psychologist*, 1952, **7**, 569–573.
6. Levy, J. Reducing the language complexity of the Study of Values: A revision. Unpublished doctoral dissertation, Univer. of Denver, 1956.
7. Odon, D. P. Flesch counts of eight current tests for introductory psychology. *Amer. Psychologist*, 1954, **9**, 143–144.
8. Stevens, S. S. & Stone, C. Psychological writing, easy and hard. *Amer. Psychologist*, 1947, **2**, 230–235: Comment 523–525.

Received July 12, 1957.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 1, 1958

THE ATTITUDES STUDENTS ASSIGN TO THEIR TEACHER

NORMAN M. CHANSKY
Oswego State Teachers College

How students see and judge their teachers are behavioral operations which have more than theoretical importance. The validity of student judgments has never been determined, yet research studies in this area tacitly assume the factor of validity. That different students viewing the same instructor assign him varied, even contradictory, attitudes warrants inquiry into the factors associated with such judging. The purpose of the present study was to examine whether attitudes assigned to a teacher can in any way differentiate between students with authoritarian outlooks and those with democratic outlooks.

PROCEDURE

After lecturing for three weeks about prenatal development, postnatal development, and theories of child development, the writer administered the Minnesota Teacher Attitude Inventory (MTAI) (**3**) to his classes in Child Psychology. The MTAI is a measure of attitude toward democratic attitudes and practices in teaching, based on the *F* Scale of the authoritarian personality series of studies (**1**). High scores on the MTAI indicate a democratic attitude; low scores, an authoritarian attitude.

After having answered the items of the MTAI, the students were asked to write on the back of their answer sheets what attitudes toward children they thought their instructor held. Care had been taken during the three introductory weeks of the course to avoid controversial issues in child psychology. At no time did the instructor consciously cue the students to his point of view.

During the subsequent 12 weeks of the course, lectures and course activities centered on (*a*) the law of effect as related to the development of behavior in children, (*b*) factors associated with acceptance of children, and (*c*) the childhood antecedents of self realization.

At the end of the course, the MTAI was readministered and the students were asked once again to state the attitude toward children their instructor holds.

RESULTS

A content analysis of the first set of student statements of the attitude which the instructor was presumed to hold yielded seven distinct categories. These categories were:

(1) freedom of children to manipulate the environment

(2) development of socially precise behaviors in children with punishment for deviants (discipline)

(3) a cold, impersonal attitude toward children

(4) the development of independence in children

(5) respect for children

(6) warm, friendly, personal relationship with children

(7) helping children with learning and emotional problems (clinical). Some attitudes could not be classified and were not included in the results.

The specific statistical tests (See Table 1) indicate that those students who felt their teacher would encourage freedom in children received significantly higher MTAI scores than any other group. On the other hand, those students who saw their instructor taking a clinical attitude toward children received significantly lower MTAI scores than most of the other groups. Their scores were not reliably different from either the discipline or the impersonal groups. In addition, the group who

TABLE 1
t RATIOS FOR THE HYPOTHECATED INSTRUCTOR ATTITUDES BEFORE EXPOSURE TO TEACHER ATTITUDE

Instructor Attitude	Number	Mean MTAI score	*t* Ratios for differences between attitude groups					
			Discipline	Impersonal	Independence	Respect	Personal	Clinical
Freedom	16	53.25	4.83[b]	2.19[a]	2.39[a]	2.27[a]	2.21[a]	4.33[b]
Discipline	12	15.42		−1.81	−5.02[c]	1.86	−3.45[b]	0.78
Impersonal	9	31.77			0.77	0.02	−0.70	1.70
Independence	12	38.04				1.04	0.12	3.50[c]
Respect	12	31.50					−0.75	3.59[c]
Personal	6	37.50						3.15[b]
Clinical	9	14.66						

Note.—degrees of freedom are $N_1 + N_2 - 2$.
[a] Significant at *P* .05 level.
[b] Significant at *P* .01 level.
[c] Significant at *P* .001 level.

looked upon their instructor as a disciplinarian received significantly lower MTAI scores than the freedom, independence, and personal groups.

These statistical tests indicate that, in the absence of well-defined cues, students projected their own attitudes toward the instructor (**2**). The reasoning behind this interpretation was as follows. Since cues to the attitude of the teacher toward children were minimized, the necessary ambiguity for projection was present. Furthermore, when those students who assigned such attitudes as freedom, independence, respect, usually associated with democracy, received higher MTAI scores, indicative of a democratic attitude, and when those students who perceived their instructor as either a disciplinarian or a clinician received lower MTAI scores, indicative of an authoritarian attitude, the interpretation of projection was confirmed. The relationship between discipline and authoritarianism is not unexpected, but why a service attitude such as that which a clinician holds received such low MTAI scores can be reconciled only if the clinical attitude represents an expression of the deep dependency needs of the authoritarian personality. If this hypothesis be substantiated in subsequent studies, it may be worthwhile investigating whether this *clinical* attitude means sensitivity to the needs of children or if it is a subtle way of expressing the wish for an anaclitic relationship with an authority figure.

Data obtained *after* exposure to the teacher's attitude are reported in Table II. Four new categories appear in the content analysis of the student assigned attitudes of the teacher. These are patience, school learning, no punishment, and understanding. A chi-square test was made to determine whether students assigned different attitudes to their instructor after he had presented his point of view. The data yielded the non-significant chi-square or 10.91 (10 *df*). Any change that took place can be attributed to chance. A coefficient of contingency of .46 was also obtained. These data further suggest that there was some consistency in responding to the teacher-attitude question on the two testings even after exposure to his attitude.

The results of the retest suggest there was greater variability than there had been in the first testing. While the freedom group had received significantly higher MTAI scores than any other group during

TABLE 2
t RATIOS FOR THE HYPOTHECATED INSTRUCTOR ATTITUDES AFTER EXPOSURE TO TEACHER ATTITUDE

Instructor Attitude	*N*	Mean MTAI score	*t* Ratios for differences between attitude groups								
			Patience	Learning	Clinical	Respect	No punishment	Discipline	Personal	Independence	Understanding
Freedom	20	56.00	0.95	2.39[a]	3.40[b]	1.05	1.16	6.17[c]	0	0.05	1.11
Patience	5	46.60		1.24	1.53	−0.16	−0.27	3.44[b]	−0.96	−0.65	−0.12
Learning	8	29.50			−0.11	−1.25	−1.52	1.50	−1.89	−1.62	−1.40
Clinical	6	31.00				−1.90	−2.33[a]	2.41[a]	−3.04[b]	−1.72	−2.02
Respect	12	47.77					−0.07	4.19[c]	−0.91	−0.60	0
No punishment	7	48.71						5.09[c]	−0.94	−0.54	−0.08
Discipline	16	10.00							−5.43[c]	−3.46[b]	−4.37[b]
Personal	14	55.92								0	0.99
Independence	5	55.40									0.57
Understanding	13	48.00									

Note.—Degrees of freedom are $N_1 + N_2 - 2$.
[a] Significant at *P* .05 level.
[b] Significant at *P* .01 level.
[c] Significant at *P* .001 level.

the pretest, the retest results indicate that the freedom group can only be distinguished from the emphasis on school learning, clinical, and the discipline groups. The emphasis on school learning is a characteristic of the authoritarian personality according to the authors of the MTAI. The results of the retest indicate further that these three authoritarian groups received significantly lower MTAI scores than many of the democratic attitude groups. The *clinical* group attained significantly lower scores than the *freedom, no punishment,* and *personal contact groups*; the *school learning* received lower scores than the *freedom* group; and the *discipline* group received significantly lower scores than all groups except the school learning. It is interesting to note that the clinical group was more democratic in attitude than the discipline group.

Again, students assigning democratic attitudes to the teacher received higher MTAI scores, indicative of a democratic attitude; students assigning antidemocratic attitudes, received lower MTAI score, indicative of an antidemocratic attitude.

SUMMARY AND CONCLUSIONS

The purpose of this study was to determine whether students' authoritarian or democratic attitudes toward children were in any way associated with their perception of their teacher.

Not having been cued to their teacher's attitude during the first phase of the experiment, the differences between those who saw their instructor as one who would encourage freedom, understand, or respect children and between those who saw him as a disciplinarian or helper of the helpless was interpreted in light of the projective hypothesis. Students, without

their awareness, assigned attitudes which they themselves held.

After having been continuously cued to the teacher's attitude thereafter, the earlier vagueness disappeared experimentally, making *projection* less tenable as an explanation, but not ruling it our entirely. The writer believes the students assigning the different attitudes to their instructor operated within different frames of reference. In this study, two frames of reference were possible: a democratic and an authoritarian. Within a democratic frame of reference, students selectively perceived in the teacher evidence for freedom of, patience with, understanding of, respect of, and no punishment of children. Within an authoritarian frame of reference, students saw their instructors as being primarily interested in discipline, school learning, and children with psychological problems.

Certain attitudes which students assign to their instructor, then, differentiate between democratic and authoritarian students. In addition, there is evidence that rating of attitudes in another person will be influenced to a great degree by the attitude the rater himself holds. Hence, there is a constant error in the judgment of democratic attitudes in others. Democratic raters are apt to give more democratic ratings; authoritarian raters are apt to give more authoritarian ratings.

REFERENCES

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N., *The authoritarian personality*. New York: Harper, 1950
2. Chansky, N. M., How students see their teacher, *Ment. Hyg.*, 1958, **42**, 118–120.
3. Cook, W. W., Leeds, C. H., & Callis, R., *Minnesota Teacher Attitude Inventory*. New York: Psychological Corporation, 1951.

Received August 1, 1957.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 1, 1958

THE SCHOOL PROGRESS AND ADJUSTMENT OF UNDERAGE AND OVERAGE STUDENTS[1]

CLYDE J. BAER

Kansas City, Missouri, Public Schools

The purpose of this study was to investigate the question of whether or not a child who begins school underage experiences similar problems and achieves the same level of development as if he had waited a year to enter school.

In this city school system a chronological age of five by November first is required for regular entrance into the kindergarten. However, children who will be five during November or December and who score a mental age of 5–0 or above on an individual intelligence test may be admitted. This study was made because some administrators and teachers in this school system feel that many of those children who enter prior to age five are too immature to be in school. However, there is no mid-year entrance or promotion and children who are not admitted must wait a full year to enter school.

PROCEDURE

Seventy-three children with birthdates in November and December were matched with 73 children with birthdates in January and February who were in the same school grade and who had entered kindergarten in September of the same year. The groups were matched on the bases of intelligence quotient, sex, and, in about two thirds of the cases, the school entered.

During their eleventh year in school the groups were compared on the bases of physical size at the time of the study, grade level attained, number of problems marked on the Science Research Associates Youth Inventory and scores on the Guilford-Zimmerman Temperament Survey. From the cumulative records, comparisons were made of marks in elementary and high school subjects, achievement test scores, teacher rating on personal traits, and number of absences.

RESULTS

Table 1 describes the groups studied.

The overage and underage students were not significantly[2] different in intelligence, as measured by the Revised Stanford-Binet, Form L, administered at time of their entrance into kindergarten or during the regular school year. For the underage students the range of IQ was from 103 to 127, and for the overage students the range was from 101 to 126.

After eleven years in school, the overage students were significantly taller but not significantly heavier than the underage students. They also had been significantly[2] more successful in maintaining regular progression from grade to grade than the underage students.

During the elementary school years (kindergarten through Grade eight) the

[1] This article is based on the writer's doctoral dissertation, the University of Kansas, 1955. Grateful acknowledgement is made to Henry P. Smith, under whose direction this work was done.

[2] Unless otherwise specified, each significant difference reported here was statistically significant at the .01 per cent level of confidence, and was obtained by the application of the t test of statistical significance, where

$$t = \frac{D}{\sqrt{\frac{N\Sigma D^2 - (\Sigma D)^2}{N^2(N-1)}}}$$

[3] In this instance the chi-square test of statistical significance was applied, and indicated a level of confidence between two and five per cent.

TABLE 1
NUMBER, IQ, AGE, HEIGHT, WEIGHT, AND GRADE OF OVERAGE AND UNDERAGE STUDENTS STUDIED

	Overage		Underage	
	Boys	Girls	Boys	Girls
Number	42	31	42	31
Mean IQ	111.17	111.31	111.24	111.40
Age				
15-5			24	10
15-6			18	21
16-3	17	21		
16-4	25	10		
Height	5'7"	5'4"	5'6"	5'3"
Weight	148	120	144	115
Grade				
8			1	
9	3		5	6
10	38	31	36	25
11	1	1		

overage students were marked significantly higher than the underage students, but the differences between the overage and underage students tended to decrease as higher grade levels were reached. The girls consistently were marked higher than the boys of their respective groups, but this difference showed no identifiable tendency either to increase or decrease. In the high school the marks received by the overage students were significantly higher than the marks made by the underage students.

TABLE 2
MEAN RATINGS ON PERSONAL TRAITS FOR OVERAGE AND UNDERAGE STUDENTS

	Underage	Overage	Difference
1. Participation in Group Activity	3.61	4.38	.77*
2. Attitude Toward School Regulations	3.84	4.30	.46*
3. Appearance	4.37	4.64	.27*
4. Dependability	3.66	4.27	.61*
5. Emotional Stability	3.71	4.70	.99*
6. Initiative	3.22	4.01	.79*
7. Cooperativeness	3.64	4.28	.64*

*Significant at the .01 % level of confidence

Achievement test scores reported at various grade levels during the elementary school showed that the overage students achieved significantly higher scores in reading for Grades three, six, and eight; in arithmetic for Grades four, six, and eight and in social studies for Grade five. The difference between overage and underage were not significant in spelling for Grade five; language for Grades five and eight; and science for Grade seven.

Near the close of each school year every student is rated by his teacher on each of seven traits. Ratings recorded for each grade level from three through eight for each trait were summed and the mean taken for each pupil. Table 2 shows the mean group ratings for overage and underage students according to a scale of one to five, with five being the highest rating. For all traits the overage students were rated significantly higher than the underage students.

The girls were rated higher than the boys of their respective groups on all traits. For three traits, Attitude Toward School Regulations, Dependability, and Emotional Stability, the differences between boys and girls were greater than the differences between underage and overage.

The number of problems marked by overage and underage students on a problem inventory was not significantly different. Although the overage and underage boys tended to mark the same problems, the underage boys marked more problems relating to home and family. The overage girls marked more problems in the category dealing with school than in all the other categories combined. Although this category was also the most popular with the underage girls, they marked almost as many problems in the section "After High School."

The results on the Guilford-Zimmerman Temperament Survey indicate that the overage boys tend to be less inclined to be suspicious or to see personal reference in

the words or actions of others, and the overage girls tend to be somewhat more socially aggressive, with greater drive and energy. All of the mean scores for the girls and all but five of the 20 mean scores for the boys fall in the average range.

Reported absences in elementary school show little difference between overage and underage students. From kindergarten through Grade eight the total median number of absences per year for both overage and underage boys was 6.50 days. For underage girls the total median was 8.00 days, and for overage girls it was 8.50 days. For both overage and underage the trend seemed to be for the greatest number of absences to occur in the primary grades, decrease in number during the intermediate grades, and then increase again at about Grade eight, with the girls showing the higher rate of increase.

Conclusions

As a group, the overage children made better school progress than did the underage children. The overage children, from kindergarten through Grade ten, made significantly higher marks in subjects, significantly higher scores on achievement tests in reading, arithmetic, and social studies, were rated significantly higher on personal traits by their teachers, and were significantly more successful in maintaining regular progression from grade level to grade level.

That the differences between boys and girls were greater than the differences between overage and underage for three of the personal trait ratings may indicate a sex-associated factor in these ratings. The overage and underage boys marked essentially the same problems on a youth inventory, but the two groups of girls did not show as much agreement on problems as did the boys.

Although there is some evidence that the differences between the overage and underage students tended to decrease with higher grade levels, perhaps this is what should be expected since the advantage in mental age that the overage group carries in the elementary school grades tends to decrease as the students get older.

Before concluding that it would be better for underage children to wait until the next year to begin school, it should be noted that most of the underage children made average school progress. As a group, they made average marks in subjects, average scores on achievement tests, received average ratings by their teachers on personal traits, and did not mark significantly more problems on the problem inventory than did the overage students.

However, it should be remembered that both the overage and the underage children studied here were selected on the basis of intelligence (average IQ of each group about 111). Thus, a better than average performance may legitimately be expected for either group on certain of the measures used.

Received August 4, 1967.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 1, 1958

A FURTHER NOTE ON BASAL METABOLISM AND ACADEMIC PERFORMANCE

MARY E. YARBROUGH
Meredith College

AND HAROLD G. MCCURDY
University of North Carolina

It is sometimes necessary to conclude that an attractive hypothesis is incorrect. In 1947 one of the present authors reported in this Journal a substantial correlation between BMR and the academic performance of a sample of college women (**2**). When his correlation was combined with the lower, but still positive correlation reported in an earlier study by Patrick and Rowles (**4**), the resulting r of .24 seemed large enough to justify the speculation that college success might owe something to basal metabolic rate. It was stressed at the time, however, that the hypothesis called for a wider sampling of the college population. In the light of the further data which we can now offer, the Meredith College sample of 1947 appears as a statistical aberration and the suggested hypothesis untenable, since the true correlation between the variables at issue seems to be in the neighborhood of zero.

NEW AND OLD DATA

In Table 1 are presented the data on which we base our rejection of the old hypothesis. It speaks for itself, but a few words regarding the studies which it summarizes will not be out of place.

The studies are listed in the table in roughly chronological order. It will be noticed that the information from Omwake, Dexter, and Lewis (**3**) was available in published form in 1934. It was unfortunately overlooked by McCurdy in 1947, but it would not have altered the suggestion made at that time. These authors state regarding their 72 Agnes Scott students: "Those making a high scholastic average tend to have high metabolism, but little relationship between poor scholarship and metabolism is evident." Their correlation of .138, when combined with the .05 of Patrick and Rowles and the .53 of McCurdy, contributes to a correlation (by Fisher's z transformation) of .19 for 152 individuals, which is significant at slightly better than the three per cent level.

The three new sets of data, all independently gathered, are those of Schutte at California, McCurdy at North Carolina, and Yarbrough at Meredith. The Schutte dissertation (**5**) has not been published, and the published abstract came to our attention quite recently; but the author has kindly furnished us with the important data which appear in the table in a personal communication.

Concerning the other two studies of recent date a few details can be given. McCurdy at the University of North Carolina at Chapel Hill, through the kindness of E. McG. Hedgpeth of the University Infirmary, secured a large collection of routine BMR records, of which a certain number applied to students. Some 560 of these were checked against the academic files of the Central Office of Records in order to find as many undergraduate students as possible who had taken courses falling within the central liberal arts curriculum. Many were excluded because their names did not appear in the files, or because they were graduate students, or because their curricula were irregular. The final yield was 117 cases. Point-hour ratio was ascertained for these

TABLE 1
SUMMARY OF STUDIES ON BMR AND COLLEGE SCHOLARSHIP

Study	*N*	BMR	Point-Hour Ratio	Correlation
Patrick & Rowles (Ohio University)	52 F	−7.3 ±8.97	1.50 ±.51	.05
Omwake, Dexter, & Lewis (Agnes Scott)	72 F			.138
McCurdy (Meredith)	28 F	−7.79 ±7.30	1.19 ±.55	.53
Schutte (University of California)	556 F	−13.08 ±9.31		−.06
	377 M	−12.36 ±10.62		.104
McCurdy (UNC)	102 (37 F, 65 M)	−8.50 ±10.35	1.13 ±.89	−.08
Yarbrough (Meredith)	33 F	−8.30 ±8.00	1.32 ±.72	−.05

Note.—BMR and point-hour ratio determinations are not exactly comparable from study to study because of differences in the machines used and in the assignment of points to letter grades.

cases in or near the academic quarter when the BMR was taken. These 117 students ranged in age from 17 to 34; the distribution was skewed, and it was decided to cut off the older students at a break in the distribution, leaving a total of 102 with an age range of from 17 to 24, symmetrically distributed around the mean age of 21. No significant differences appeared between the 65 men and 37 women of this sample in age, point-hour ratio, BMR, or the relations between variables, and therefore the results are lumped together in the table. It was fully realized at the time that the manner of assembling these data did not afford a clean-cut test of the hypothesis, since errors might have crept in at several points; but the specific purpose was to discover whether the hypothesized relationship was strong enough to stand up in the face of the sort of random errors which might be encountered by a practical-minded college dean or physician working upon the same hypothesis. The physician who lent the BMR records, incidentally, felt that the hypothesis was perhaps a realistic one.

Yarbrough at Meredith, on the other hand, attempted to follow strict criteria both in sampling the students and in taking the BMR readings. The students were all in their first year of college, taking virtually the same program, and beir subjected to very similar requirements by teachers who value good academic work. Three BMR records were taken on each girl, and each record was carefully inspected for possible technical flaws. As in the earlier Meredith study, the record yielding the lowest BMR for a given individual was utilized in the correlational analysis. In all essential respects the experimental procedures and the nature of the sample seemed comparable to those of the earlier study. But the correlation between BMR and scholarship was definitely much lower.

It is clear from inspection of Table 1 that the correlation of .53 in 1947 has to be considered as quite exceptional. We can think of no explanation except accident of sampling. The total weight of the evidence supports the view that BMR's in the normal range have little or nothing to do with scholarship at the college level.

Conclusion and Commentary

The available evidence contradicts the plausible hypothesis that basal metabolic rate might affect academic performance in college. The correlation between the physiological and the intellectual indices used appears to be at or near zero. Perhaps the persuasiveness of the hypothesis in the first placed depended entirely too much on a superficial view of the interrelationships of energy, motivation, and thought.

In psychological circles the null hypothesis is often regarded with distaste, and "negative" results are dreaded so much that, it is rumored, some investigators avoid repeating experiments lest chance should be against them the second time. If the rumor is true, then some of our theories must have very spindly foundations. No matter what the significance level of a single result may be, it is not sufficient by itself to establish a general working rule. The present summary of metabolic studies should make it clear that one correlation in a set of six may be freakishly different from the rest.

REFERENCES

1. Fisher, R. A. *Statistical methods for research workers. Edinburgh:* Oliver & Boyd, 1941.
2. McCurdy, H. G. Basal metabolism and academic performance in a sample of college women. *J. educ. Psychol.*, 1947, **38,** 363–372.
3. Omwake, K. T., Dexter, E. S., & Lewis, L. W. The interrelations of certain physiological measurements and aspects of personality. *Char. & Person.*, 1934, **3,** 64–71.
4. Patrick, J. R., & Rowles, E. Intercorrelations among metabolic rate, vital capacity, blood pressure, intelligence, scholarship, personality and other measures on university women. *J. appl. Psychol.*, 1933, **17,** 507–521.
5. Schutte, H. M. An investigation of basal metabolic rate and college scholarship. Unpublished doctoral dissertation, Univer. of California at Berkeley, 1951.

Received August 13, 1957.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 1, 1958

AN EXPERIMENTAL PROGNOSTIC TEST FOR REMEDIAL READERS

RICHARD W. WOODCOCK
Oregon College of Education

A remedial reading program is usually a relatively expensive program since it may entail individualized or small group instruction in addition to the school's regular full-time program. Limitations of finances and time usually limit the amount of service available and it is desirable to utilize some sort of selective procedure in choosing cases for remedial instruction.

This article describes the development and evaluation of a test for selecting the remedial readers most likely to profit from special instruction. For the purposes of this discussion, a "remedial reader" is defined as any child whose reading achievement is significantly retarded below both his grade placement *and* his capacity or potential for reading achievement. In addition, the child in mind is specifically lacking in reading skills usually developed by the third or fourth grade.

The test is designed to duplicate as nearly as possible the "learning-to-read" process for remedial readers. It was assumed that the primary learning task of the remedial reader is to learn associations between written symbols and familiar spoken language. By means of the experimental test an attempt has been made to provide the remedial reader with a controlled learning situation uniquely different from reading, yet so similar to the process of learning-to-read that performance on the test might be indicative of ability to profit from remedial reading instruction.

DESCRIPTION OF TEST

The experimental test is basically composed of five short stories written entirely with pictures and symbols instead of words. Each of these "test stories" is on a successively more difficult level and is preceded by a practice page which introduces the new symbols used at that level and provides practice with the symbols in context. The manner of presenting the material on these introductory pages is similar to that of a reading lesson. Figure 1*a* shows the first group of symbols and the following line of practice material for Level I. A portion of the Level V test story is shown in Figure 1*b*.

The total vocabulary of the experimental test consists of 72 symbols. Two of these symbols represent the inflections, "-ing" and "-s". The "-s" inflection is used both as a plural and to denote possession.

The vocabulary of the first level consists of 14 symbols. These symbols represent such words as "boy," "the," "cat," "is," "black," and so forth. Many of the symbols at the first level consist of stick drawings or simple pictures of the object the symbol is intended to represent. Each successive level adds approximately 15 new symbols to the vocabulary being used in the test stories.

The symbols at the higher levels do not provide picture clues. There are some phonetic elements common to a number of the symbols. However, this is not pointed out to the *S*s.

The materials for administering the test consist of an answer sheet and a test booklet. The test booklet contains ten pages, two pages for each of the five levels, as described above. The answer sheet is used by the examiner in recording the errors made by the *S* while reading the test stories.

The score on the test is the total

Fig. 1. (*a*) First Two Lines From the Level I Introductory Page (Boy, Girl, Dog, Cat, And—Boy and Girl, Cat and Girl, Boy and Dog) (*b*) First Two Lines From The Level V Test Story (Mother went to a surprise party. So did the boy and girl. It was a little girl's birthday so they took presents for her.)

number of errors made on the five test stories read orally by the *S*. Hesitations of more than approximately five seconds (for which the word is supplied by the examiner), substitutions, omissions, and additions were considered errors. Repetitions were not so counted in this study.

Procedure of the Study

1. The first step in the development of the instrument consisted of small-scale tryouts and revisions with a series of crude tests and procedures. From this work evolved the Experimental Prognostic Test for Remedial Readers described above.

2. The experimental test was administered to a complete fourth grade class of 24 pupils at Francis Willard School in Eugene, Oregon during March of 1955. The resulting data were analyzed for reliability and concurrent validity information.

3. Next the test was administered to 26 remedial readers in the Corvallis Public Schools for the purpose of evaluating the predictive validity. Each of these 26 remedial readers was given the experimental test and a series of Gates reading tests. Following a four-month period of daily individual and small-group instruction these same children were given an alternate form of the reading tests. These data were analyzed for predictive validity information.

TABLE 1
Correlation of Experimental Test Error Scores with Gates Reading Tests
(N = 24)

Vocabulary	−.51
Comprehension	−.65
Speed	−.58
Oral reading	−.48
Composite score	−.59

Findings

Concurrent validity. Twenty-four children in a regular fourth-grade classroom were administered the Gates Reading Survey, the Gates Oral Reading Test, and the Experimental Prognostic Test. The experimental test was found to correlate −.59 with the composite measure of reading achievement. (Negative correlations are based on error scores.) Table 1 presents the correlations of the experimental

test with the four reading achievement subtests. These correlations are all significant at the .05 level of confidence.

A mean score of 58.0 errors was made on the experimental test by this fourth-grade class. The standard deviation of scores was found to be 22.8 errors.

Reliability. The experimental test reliability, as computed by the Kuder-Richardson Formula 21, is .91. This value probably represents an underestimate since the data do not meet the assumption of equal difficulty for all items in the test. The split-half correlation was .92 and a reliability of .96 is obtained when this value is corrected by the Spearman-Brown Formula. These values are probably overestimated to a small degree. The standard error of measurement, based on the corrected split-half reliability, is 4.6 errors. Since the standard deviation of scores, in this sample, is 22.8 errors it would seem to indicate that the test may have good discriminative properties.

Predictive validity. Estimates of the predictive value of the experimental test were obtained from a group of cases referred and selected for remedial reading instruction. Since the total number of cases was relatively small and spread among several age levels, it was not possible to apply correlational procedures to the data which had been collected. An alternative method was developed which consisted of pairing the remedial cases and predicting, on the basis of the experimental test scores, which member of each pair would show the most gain by the end of the training period. In order to pair the cases they were grouped according to age levels and each child was paired with each other child at the same age level if there was a significant difference between their scores on the Experimental Prognostic Test. Several different values were established as "significant differences" in order to observe the amount of variance in the predictive ability of the test. These significant differences in scores were arbitrarily established as two, three, five, eight, and ten times the standard error of measurement. On this basis it was usually possible to pair each child with a few other children at the same age level. The member of each pair showing the better performance on the prognostic test was predicted as the one most likely to gain from the special instruction.

This procedure experimentally duplicates a very real and practical administrative problem in the area of remedial reading. The person responsible for selecting remedial reading cases is faced constantly with the question, "Which of these two children can profit the most from remedial instruction?" In essence, this study has experimentally created a series of such possible choice situations. The statistical purpose of this portion of the study was to determine whether the proposed instrument would predict, any better than chance, which children would profit the most from instruction.

Table 2 shows the results of this portion of the study. Using two times the standard error ($2 \times s_e$) as a significant difference between prognostic test scores, it was possible to set up 44 pairs of remedial cases. Of these 44 predictions, 34 were correct—indicating a predictive efficiency of .54 (54% better than chance).

TABLE 2

PREDICTIVE EFFICIENCIES OF EXPERIMENTAL PROGNOSTIC TEST WITH CASES GROUPED BY CHRONOLOGICAL AGE ($N = 26$)

Differential (s_e units)	Pairs (*n*)	Correct predictions	Percentage correct	Predictive efficiency
2X	44	34	77	.54
3X	39	31	80	.60
5X	25	20½	82	.64
8X	17	15	88	.76
10X	12	11	92	.86

When higher values are used as a significant difference between scores, the number of pairs becomes less and the accuracy of prediction steadily increases. Using a value of ten times the standard error provides 12 pairs and a predictive efficiency of .86. The steady increase in predictive efficiency, as higher values are used for discrimination, is further evidence that the test is functioning predictively in this situation.

When these same cases are paired on the basis of grade level rather than age level it is possible to obtain more pairings, but there is a decrease in the predictive efficiency of the instrument. This is probably due to the more heterogenous nature of the sample in terms of age. (At one time the experimental test was administered to a number of children at various ages, and it was found that the number of errors rapidly decreases with an increase in the age of the *Ss*.) If a value of five times the standard error is used as a significant difference between scores it is possible to set up 35 pairs. This yields a predictive efficiency of .60 (as compared to .64 for 25 pairs when grouped by age). A predictive efficiency of .60 was the highest obtained when pairings were made on the basis of grouping by grade (as compared to .86 when grouped by age).

Twenty-three of the cases used in this portion of the study had had Stanford-Binet Intelligence Tests administered at some time. Predictive efficiencies were computed for these data using the same procedure used to predict with the experimental test. No predictive value for the Stanford-Binet was found in this sample. However, since this portion of the study was subjected to limited control the results do need to be interpreted cautiously. One significant point is brought out by this comparison. It was quite difficult to obtain as large a number of pairings from the intelligence test data as from the experimental test data. This was due to the relatively narrow range of performance available for discrimination between *Ss* on the intelligence test results as compared to the experimental test results. The range of I.Q.'s was approximately 30 points while the range of performance on the experimental test was approximately 100 points for the same *Ss*. In both instances, quite coincidentally, the size of the standard error of measurement is nearly identical.

Conclusions

Within the limitations of this study it would appear that the Experimental Prognostic Test for Remedial Readers has demonstrated a predictive value in selecting cases for remedial reading instruction. Further research will be required in order definitely to determine the degree of this value. At present, it would appear that the test shows a greater possibility for *predicting* success based on gains during a period of instruction than the techniques in common use. However, it should be noted that the research concerning *the predictive value* of the techniques in current use is incomplete.

One is impressed by the paucity of research studies designed to consider some aspect of the prediction of success in remedial reading. Most studies which are referred to in this area have been originally designed for some other purpose. For example, the writing and research which provides the basis for the use of intelligence tests as a predictor of success in remedial reading is primarily based on studies of concurrent validity obtained between intelligence and reading achievement with groups of normals. These correlations have not been based on the

predictive validity between the intelligence test scores of retarded readers and their subsequent gains in remedial reading. This does not mean to say that the studies were not appropriate for the purposes originally intended by the authors, but the application of the results of such studies to the problem of selecting cases or predicting success in remedial reading will remain highly questionable until subjected to experimental verification.

Received August 30, 1957.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 1, 1958

EFFECT OF AN AUDIO-VISUAL PHONICS AID IN THE INTERMEDIATE GRADES

CAROLYN LUSER, EILEEN STANTON, AND CHARLES I. DOYLE
Loyola University, Chicago

To test the effect of formal phonics drill on a group basis, the present experiment employed audio-visual aids consisting of uniform recordings and individual charts for each pupil in four experimental rooms.[1]

The population was chosen from a lower socioeconomic area, with many semitransients. The sample contained many Negroes, Puerto Ricans, Mexicans, and a scattering of other nationalities. The area and population were selected in the hope of finding more than the average number of handicapped readers. This hope was realized when the pretests were scored. The average reading scores in all eight rooms were below the grade-level expectancy, and there were many seriously retarded readers. The number of drop-outs and absences confirmed the transient nature of the population. While nearly 300 pupils were available for pretests, only 214 completed the entire battery of tests at the conclusion of the experiment.

[1] Permission for the use of these four schools was graciously given by Peter B. Ritzma, Chairman, Special Projects Committee, and Don C. Rogers, Assistant Superintendent, Chicago Public Schools, and by Rev. David C. Fullmer, Assistant Superintendent, Chicago Catholic School Board. The District Superintendent, Douglas Van Bramer, and the principals and teachers in the four schools also rendered much cooperation.

The authors are indebted to the World Book Company for permission to reprint Stanford Primary Reading Test, Form D. It appeared to the authors that fill-in responses would be more valuable for detailed analysis than the multiple-choice style adopted in the 1953 forms.

The phonographs, phonics records, and pupil charts for the experimental rooms were donated by Bremner-Davis Phonics.

Two public and two parish schools were available in the area selected. The grade levels chosen for the experiment were the third and fourth grades, where reading difficulties tend to become more evident, and where pupils' competence to follow group test instructions may be assumed more safely than at lower levels.

In each school an experimental room and a control room were selected. The criterion for the selection of the experimental room was the lower average IQ derived from the Kuhlmann-Anderson Test, Form D (sixth edition). This form not only fitted the average grade expectancy of the experimental and the control rooms, but had the further advantage that the content was about equally divided between verbal and nonverbal subtests.

Besides the intelligence tests, four achievement measures were secured for each pupil; Gray's Oral Reading Paragraphs, Stanford Primary Reading, Form D (paragraph meaning and word meaning) and the Marion Monroe form (written) of the Ayres Spelling Scale (**3**).

After the pretests, each of the four experimental rooms received 43 twenty-minute sessions of phonics drill with the phonograph records and individual pupil charts (**1**). These sessions were spaced three times a week for a period of 15 weeks. The experiment was limited to 15 weeks because the pretests had to wait for mid-year promotions, and the retests had to be completed before final examinations in June. No special motivation was given during the drill periods, aside from the encouragement offered in the records themselves. All the phonics sesions were conducted by one of the workers who had

administered the pretests, who was thus a familiar figure to the children.

At the conclusion of the experiment, the entire population, experimental and control, was retested with the complete battery described above. Retests were conducted by the same examiners who administered the original tests.

Results

The data from both batteries were analyzed by two methods: by computation of the standard error of the gains (**2**), and by chi square. The chi-square analysis was based on the assumption that a gain in individual scores in the experimental group substantially greater than the average gain evidenced by the control group would be valid to determine a cutting point. Accordingly, for the achievement tests, a gain of .5 grade score in 3.5 months was set as significant. For the K-A IQ, allowing for practice effect much greater than that shown by the control group, a gain of three IQ points was chosen.

While the chi-square test was a less refined and precise statistic than that based on standard error, it did in general corroborate the findings of the latter. The results of both methods are summarized in Table 1.

The gains in three of the four achievement scores seem to indicate clearly the effectiveness of the standardized audio-visual drill. Failure to show a similar gain in word meaning is understandable in view of the background of the children in this sample. A marked gain in word meaning would seem to depend on enriched experience and resultant growth in vocabulary more than on mastery of phonics.

It seems evident that the unexpected gain in IQ points does not indicate a real change in intelligence. It does serve to point up the fact that performance on a group test of intelligence is markedly dependent on the pupils' familiarity with the printed word. It is suggested, however, that improved habits of attention from group drill sessions may have contributed

TABLE 1
Test Scores of Experimental and Control Groups Before and After Audio-Visual Phonics Drill

Tests		Experimental		Control		Net Diff. in gain	t	P	Chi square	P
		Pre	Post	Pre	Post					
		N = 105		N = 109						
Oral Reading	M	3.0	3.76	2.97	3.27	.46	3.65	.001	13.93	>.001
	σ	1.15	1.40	1.13	1.29					
Parag. Mean.	M	2.84	3.61	2.90	3.30	.37	4.40	>.001	8.57	.01
	σ	.84	1.16	.92	.97					
Word Mean.	M	2.79	3.26	2.76	3.16	.07	1.11	.30*	1.22	.30*
	σ	.78	.93	.83	.89					
Spell.	M	2.79	3.26	2.76	3.00	.23	2.35	.02	4.26	.05
	σ	.97	1.17	.84	.94					
IQ	M	88.3	94.5	92.3	93.0	5.50	4.91	>.001	23.28	>.001
	σ	11.2	11.9	11.0	13.0					

* Not significant

in part to more successful performance on the K-A retests of the experimental group.

Summary

A sample of 214 third and fourth graders from four schools in an underprivileged area was divided into two groups. A control group was measured against an experimental group which received 43 drill sessions in phonics, with uniform recordings and individual charts. When both groups were retested after the experiment, the experimental group showed gains on standard tests of oral reading, paragraph meaning, and spelling, which were significantly greater than the gains of the control group. Gain in word meaning apart from context was not significant. Retest performance on a paper-and-pencil intelligence test also showed a marked gain for the experimental group.

REFERENCES

1. Bremner, A. J., & Davis, Josephine. *The sound way to easy reading.* Wilmette, Ill.: Bremner-Davis Phonics, 1953.
2. McNemar, Q. *Psychological statistics,* New York: J. Wiley, 1949.
3. Monroe, M. *Children who cannot read,* Chicago: Univer. of Chicago Press, 1932.

Received September 19, 1957.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 1, 1958

PREDICTING SCHOOL ACHIEVEMENT FOR BILINGUAL PUPILS[1]

JAMES G. COOPER
Territorial College, Agana, Guam, Marianas Islands

Teachers and other school personnel of the Territory of Guam have long been aware of the need for adequate predictors of school achievement for their bilingual pupils. In December of 1956, Guam's Department of Education authorized a study to discover to what extent, if any, currently available tests would provide such predictions. Guam is a territory whose cultural patterns are rapidly changing. From a Spanish type of church-dominated society, from a military paternalism, Guam is becoming infused with current American ideas, trends, and practices. The local language, Chamorros, is slowly giving way to English. However, English is spoken only in the school classroom, infrequently on the playground, and rarely in the home and community. Consequently, because of the unique cultural and language factors, currently available measures of intelligence must lie in the realm of questionable validity until demonstrated otherwise.

This study endeavored to determine the predictive ability of six tests of intelligence for certain fifth-grade pupils of Guam. Only those tests which were wholly or partially performance or nonverbal were considered. In order to hold cultural factors constant, four relatively isolated communities were selected. The villages of Inarajan, Merizo, Talofofo, and Umatac have had no electricity (and consequently no television) until quite recently; no statesiders (persons whose usual abode is mainland, U.S.A.) as residents; no telephones; few movies. Books and magazines are not commonly found. These four villages enrolled a total of approximately 180 fifth-grade children distributed among six classes.

The plan of study entailed the following:

1. Administer three group tests of intelligence to all fifth-grade pupils of the four villages. The tests were the California Tests of Mental Maturity, 1950 S-Form, Elementary; Davis-Eells Games, Intermediate Level; and the Culture Free Intelligence Test, Scale 2, Form A.
2. Select a stratified, random sample of 51 pupils from the larger group. Give each of these pupils the following *individual* tests: Leiter International Performance Scale, Wechsler Intelligence Scale for Children, and the Columbia Mental Maturity Scale.
3. Give all pupils the California Achievement Tests, Form AA, Elementary level.
4. Obtain teacher ratings for each child regarding his school success.
5. Relate intelligence tests scores to achievement test scores and to teachers' ratings (applicable to group tests only).
6. Study the interrelationships between certain tests.

PROCEDURE

The plan indicated above was followed: The Davis-Eells Games and California Tests of Mental Maturity were administered between November 1956 and February 1957. The Culture Free Intelligence Test was given during March of 1957 and the California Achievement Tests given in May 1957. The teachers' ratings were obtained prior to achievement testing. Individual tests were given from February through June 1957 in this sequence: Wechsler Intelligence Scale for

[1] A research supported by a grant from the James McKeen Cattell Fund.

Children, Columbia Mental Maturity Scale, and Leiter International Performance Scale.

The sample for individual testing was drawn via a table of random numbers. Each of the six classrooms was permitted to contribute its proper share of boys and girls. This was necessary because the schools of Inarajan and Merizo divided their fifth-grade pupils into fast and slow groups. Also, these schools enrolled twice as many fifth graders as Umatac and Talofofo.

Group Tests

The principal findings from the three group tests of mental ability are shown in Table 1. The table shows that 164 pupils obtained a mean California Test of Mental Maturity total IQ of 83.494 with a standard deviation of 11.087; the Pearson correlation coefficient between these scores and the California Achievement Tests was .644 and the stability of this coefficient is indicated by the .99 confidence limits of .509–.747. These limits were computed via the z transformation (**6,** p. 147). Similar data are recorded for California Language IQ, Nonlanguage IQ, Davis-Eells Games, and Culture Free Intelligence Test. The latter test is reported in raw scores because the IQs yielded a distribution severely skewed, i.e., too many low scores. The use of raw scores gave a normal (by eye) distribution.

A positive skew was noticed in the distribution of California Nonlanguage IQs. However, a chi-square test indicated that the obtained frequency distribution fitted the normal form adequately (**3,** pp. 284–285).

The Davis-Eells Games scores were considerably lower than the other two tests. The possibility existed that this may have been caused by the pupils' difficulties in understanding the directions. (This test consists of a number of pictures about which the examiner makes various statements. The subject responds by indicating which statement best applies to each picture.) This possibility was explored by retesting a fifth-grade class in Merizo. The second test was given

TABLE 1

CORRELATION COEFFICIENTS BETWEEN GROUP TESTS OF INTELLIGENCE AND CALIFORNIA ACHIEVEMENT TEST RAW SCORES

Group Test	M	σ	r	.99 confidence limits (via z)	N
California Test of Mental Maturity					
Total IQ	83.494	11.087	.644	.509–.747	164
Language IQ	81.024	10.194	.584	.434–.703	164
Nonlanguage IQ	88.067	17.239	.522	.359–.655	164
Davis-Eells Games (IPSA)	66.970	11.146	.531	.369–.601	164
Culture Free Intelligence Test (Raw scores) (this yields a mean IQ between 75 and 78)	20.685	6.355	.549	.391–.676	163
California Achievement Test					
Total raw score	154.451	36.178			164

wholly in Chamorros. The average increase of three points lacked statistical significance. It was concluded that this test measures as well in English as it does in Chamorros.

The matter of sex differences was analyzed for each of the group tests. None of the small differences approached the 5% level of significance. The mean CA of the boys was 12–0 and of the girls 11–8; the difference of four months was significant at 1%.

Discussion of group tests. The data of Table 1 indicate that the California total IQ predicted California Achievement Test scores fairly well; the schools of Guam should consider seriously more widespread application of this test. It was interesting to note that neither the Language nor Nonlanguage IQ's should be used separately.

Neither the Davis-Eells Games nor the Culture Free Intelligence Test offer as much promise. However, from the point of view of test theory, the fact that these two tests do show pronounced positive correlations with scores on a typical school achievement test is highly significant. In other words, the abilities sampled by these two measures are those which in part determine school success in a bilingual setting.

Group tests and teachers' ratings. The six teachers of the children tested were asked to rate their pupils according to the directions: "... Divide your pupils into three groups. Place the names of your best pupils into the top third, the poorest into the low third, and the others into the middle third." The relationship between this sort and the four group tests was established by locating the median for each class and preparing a 3 x 2 table from which chi square could be computed. A high value of chi square indicates that the teachers' judgments and the pupils' test positions (above or below the medians for their classes) were congruent. These data appear in Table 2.

TABLE 2

CHI-SQUARE VALUES RESULTING FROM COMPARING TEACHERS' RATINGS WITH TESTS OF ABILITY AND ACHIEVEMENT

(df = 2)

School	N	Name of Test					
		Calif. Mental Maturity			Davis Eells	Culture Free	Calif. Achievement
		Total	Language	Nonlanguage			
Inarajan							
Teacher A	29	9.48**	1.73	11.59**	2.82	5.93	19.08**
Teacher B	28	7.42*	4.47	6.36*	2.74	2.84	10.11**
Merizo							
Teacher A	23	1.30	1.92	1.81	1.60	4.71	6.19*
Teacher B	32	1.28	4.70	3.61	8.06*	2.44	16.02**
Talofofo	34	4.49	7.63*	3.52	4.77	1.07	12.77**
Umatac	26	7.26*	3.22	9.96**	1.96	10.95**	17.91**

* Significant at 5% level.
** Significant at 1% level.

The table shows that the California Test of Mental Maturity IQ's agreed fairly well with the ratings of three teachers and that the Nonlanguage IQ was as effective as the Total IQ. This finding was surprising as teachers may be expected to give more weight to their pupils' verbal behavior than to their nonverbal behavior. Neither the Culture Free Intelligence Test nor the Davis-Eells Games corresponded well with teachers' ratings. The California Achievement Test, however, showed a pronounced agreement with teachers' ratings. These latter data demonstate a high degree of validity for this test, i.e., the California Achievement Test seems to measure the kinds of achievements that these teachers deem important when they differentiate between successful and less successful pupils. These data also suggest that this test may possess considerable curricular validity. This point, however, should be verified by study of both teachers' opinions and actual curricular materials used in these classes.

Individual Tests

The results obtained from administering the three individual tests of intelligence and correlating these scores with the total raw scores from the California Achievement Tests appear in Table 3. The table shows that the best prediction was given by the Verbal Scale of the Wechsler Intelligence Scale for Children. The greatest amount of scatter was given by the Columbia Mental Maturity Scale as shown by its sigma of 24.37; this test also produced the highest mean IQ.

The means of these individual tests were well below those given in the respective manuals. The obtained sigmas were three to four IQ points lower than those reported for the Wechsler Intelligence Scale for Children (**8**), equal to the Columbia Mental Maturity Scale's listed 25 IQ points (**1**), and lower than the figures given for the Leiter International Performance Scale (**4**). The latter was difficult to determine, because the Leiter manual lacks specificity on this point. These data indicate that the IQ scores for the Guam sample were distributed in a manner similar to the standardization groups.

Discussion of individual tests. The data revealed in Table 3 show that the Verbal Scale of the Wechsler Intelligence Scale for Children gives a fairly accurate pre-

TABLE 3

PEARSON CORRELATION COEFFICIENTS BETWEEN CALIFORNIA ACHIEVEMENT TESTS, (RAW SCORES) AND IQS FROM INDIVIDUAL TESTS

(N = 51)

Test	M	σ	r	.99 confidence limits (via z)
Wechsler Intelligence Scale for Children				
Full Scale	72.89	11.84	.77	.58–.88
Verbal Scale	71.58	11.30	.80	.63–.90
Performance Scale	77.15	12.70	.54	.24–.75
Columbia Mental Maturity Scale	83.86	24.37	.61	.34–.79
Leiter International Performance Scale	72.78	12.58	.66	.40–.82
California Achievement Test (raw score, sum of all tests)	155.39	43.16		

diction of school achievement in Guam's bilingual setting. It was interesting to note that the Wechsler Full Scale IQ was not quite so efficient a predictor as was the Verbal Scale IQ. This finding was not supported by comparable evidence concerning the California Test of Mental Maturity reported in Table 1; in the latter case, the Language IQ and the Nonlanguage IQ predicted about equally well. Wechsler states that of the Verbal Scale subtests, "Information" and "Comprehension" are poor for those with inadequate verbal facility and that "Arithmetic" is influenced by education (**7**, pp. 80–82). He also notes that "Vocabulary" is affected by schooling (**7**, pp. 98–99). Therefore, it seems reasonable to believe that in a bilingual setting, successful school achievers will obtain higher scores on both the Wechsler Verbal Scale and on the California Achievement Test, i.e., we may be dealing with common elements rather than with intelligence per se.

The data relating to the Leiter International Performance Scale are of considerable concern. This test is expensive (about $200), time-consuming to administer (from 45 to 90 minutes), and very bulky. Its correlation with the California Achievement Test, although high (r equaled .66), was only slightly higher than the relatively inexpensive Columbia Mental Maturity Scale. Further, the Columbia may be administered in from 10 to 15 minutes, a factor of real importance to many psychometrists. On the positive side, however, both the Leiter and the Columbia have demonstrated considerable validity for use with these bilingual pupils. These data are all the more significant when it is recalled that the Leiter is a completely nonverbal test and the Columbia is almost so (five cards on the Columbia require the ability to read words or letters, and five involve numerals).

Interrelationships between tests. The Leiter International Performance Scale and the Columbia Mental Maturity Scale are relatively new arrivals to the scene

TABLE 4

INTERCORRELATIONS BETWEEN TESTS

(N = 51)

Test	Leiter International Performance Scale	Columbia Mental Maturity Scale
California Test of Mental Maturity		
Total IQ	.68	.62
Language IQ	.62	.54
Nonlanguage IQ	.66	.60
Wechsler Intelligence Scale for Children		
Full Scale IQ	.83	.74
Verbal Scale IQ	.73	.66
Performance Scale IQ	.78	.68
Davis-Eells I.P.S.A.[a]	.72	.69
Culture Free Intelligence Test, Raw scores (N = 50)	.75	.60
Columbia Mental Maturity Scale IQ	.69	

[a] Index of Problem Solving Ability.

of mental measurements. In order to clarify the nature of their functioning, their scores were correlated with each other and with the other tests of mental ability utilized in this study. The results are shown in Table 4. The table indicates that the Leiter measures quite consistantly with the other tests, since the correlation coefficients ranged from .62 (California Test of Mental Maturity, Language IQ) to .83 (Wechsler Full Scale IQ). The Columbia followed a pattern similar to the Leiter, but the coefficients were somewhat lower.

Summary and Conclusions

This study was undertaken to ascertain to what degree, if any, currently available measures of intelligence predict school achievement for the bilingual pupils in the Territory of Guam. Three group tests, the California Test of Mental Maturity, 1950 S-Form, Elementary; the Davis-Eells Games, Intermediate Level; and the Culture Free Intelligence Test, Scale 2, Form A were given to 164 pupils in grade five. Three individual tests of intelligence: the Leiter International Performance Scale, the Wechsler Intelligence Scale for Children, and the Columbia Mental Maturity Scale were given to a stratified, random sample of 51 pupils. School achievement was defined primarily by scores received on the California Achievement Tests, Form AA, Elementary Level, and secondarily by teachers' ratings.

All the intelligence tests correlated positively with the California Achievement Tests. The correlation coefficients ranged from .53 to .77 as follows: Davis-Eells Games, .53; Culture Free Intelligence Test, .55; Columbia Mental Maturity Scale, .61; California Tests of Mental Maturity, .64; Leiter International Performance Scale, .66; and the Wechsler Intelligence Scale for Children, Full Scale, .77.

Although teachers' ratings corresponded well with rank on the achievement test, they were not closely related to scores on the group intelligence tests.

This study demonstrated that the six intelligence tests examined predicted school success with a degree of accuracy ranging from moderate to high for Guam's bilingual pupils.

REFERENCES

1. Burgemeister, B. B., Blum, L. H., & Lorge, I. *Manual, Columbia Mental Maturity Scale.* Yonkers-on-Hudson, N. Y.: World, 1954.
2. Cattell, R. B., & Cattell, A. K. S. *Handbook for the Culture Free Intelligence Test, Scale 2.* Champaign, Ill.: Institute for Personality and Ability Testing, no date.
3. Guilford, J. P. *Fundamental Statistics in psychology and education,* (2nd ed.) N. Y.: McGraw-Hill, 1950.
4. Leiter, R. G. *Part I of the manual for the 1948 revision of the Leiter International Performance Scale.* Chicago: C. H. Stoelting, no date.
5. Lindquist, E. F. *Statistical analysis in educational research.* N. Y.: Houghton Mifflin, 1940.
6. McNemar, Q. *Psychological statistics,* (2nd ed.) N. Y.: Wiley, 1955.
7. Wechsler, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.
8. Wechsler, D. *Wechsler Intelligence Scale for Children; Manual.* N. Y.: Psychological Corp., 1949.

Received October 1, 1957.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 1, 1958

THE EFFECT OF CHILD PSYCHOLOGY ON ATTITUDES TOWARD PARENT-CHILD RELATIONSHIPS[1]

FRANK COSTIN

University of Illinois

While most psychologists who teach in our colleges and universities are very much interested in changing their students' attitudes, few of them are reporting studies of the problem. In their review of research in the teaching of psychology, Birney and McKeachie (**1**) show the paucity of this kind of investigation. Furthermore, most of the attitude studies which they describe deal with the introductory course. Little is being done to discover the nature of attitudes related to advanced courses in psychology, either on the graduate or undergraduate level. The study presented in this paper is intended to help close that gap.

A second-level course which is becoming increasingly popular with undergraduates is child psychology. The students who take this course vary considerably in their educational and vocational goals, but almost all of them share at least one common objective: they hope to learn something about children which will help them as future parents. They are especially interested in learning how their own attitudes might affect their children's development. It is important, therefore, that those who teach this course find out what kinds of attitudes toward parent-child relationships their students have, and how these attitudes change when they study child psychology. As a step in this direction, a recent pilot study by Hurley and Laffey (**2**), involving 19 students, concluded that a 10 week's child psychology course at Michigan State University succeeded in making these students less rejecting in their attitudes toward children. There was no change in overprotecting attitudes.

The present paper reports a more extensive investigation of the same kind of problem explored by Hurley and Laffey. It attempts to answer this question: To what extent can an undergraduate, one-semester course in child psychology change students' attitudes toward parent-child relationships?

METHOD

The subjects of this study consisted of four different classes totaling 157 students. None of them had previously taken a course in either child psychology, child development, or family relations. They represented a variety of majors within the College of Liberal Arts and Sciences at the University of Illinois, and were also drawn from other colleges within the university.

The content of the course was the same for each class, with emphasis on problems of parent-child interaction. All classes were taught by the same instructor. In addition to the use of a basic text and lectures as sources of information and attitudes, films were shown and discussed. Supplementary readings also helped to extend the variety of course content.

A parent-child attitude scale was administered to each class at the beginning and end of the course. The scale was also given at the beginning and end of the semester to 155 undergraduates at the same university who were enrolled in a one semester introductory course in sociology. This was done to see whether or not changes which might occur in the psychol-

[1] A briefer version of this paper was read on September 4, 1957, at the convention of the American Psychological Association in New York City.

ogy students could be simply the result of general college living rather than taking child psychology. Like the psychology classes, the sociology students represented a variety of majors within the College of Liberal Arts as well as other colleges. They were approximately the same as the psychology students in age and college class. None had ever taken a course in child psychology, child development, or family relations, nor was any taking one at the times the scale was administered.

The attitude scale was a slightly modified version of one constructed and validated by E. J. Shoben (**4**). Ten items from Shoben's scale were omitted, because they represented a miscellaneous group which lacked homogeneity, and therefore were not as meaningful for this study as the other items. In all other respects the scale was identical with Shoben's. It consisted of 75 statements about parent-child relationships to which students responded by strongly agreeing, mildly agreeing, mildly disagreeing, or strongly disagreeing. The scale measured three kinds of attitudes: *dominating* (40 items), *possessive* (20 items), and *ignoring* (15 items).

Dominating attitudes "reflect the tendency on the part of the parent to put the child in a subordinate role, to take him into account quite fully but always as one who should conform completely to parental wishes under penalty of severe punishment" (**4**, p. 137). The following statements from the scale represent this kind of attitude: "A child should have strict discipline in order to develop a fine, strong character." "It is sometimes necessary for the parent to break the child's will."

Possessive attitudes "reflect the tendency to 'baby' the child, to emphasize unduly (from a mental hygiene viewpoint) the affectional bonds between parent and child, to value highly the child's dependence on the parent, and to restrict the child's activities to those which can be carried on in his family group" (**4**, p. 137). Statements illustrating this attitude are: "The best child is one who shows lots of affection for his mother." "Children should always be loyal to their parents above anyone else."

Ignoring attitudes reflect "the tendency on the part of the parent to disregard the child as an individual member of the family, to regard the 'good' child as one who demands the least parental time, and to disclaim responsibility for the child's behavior" (**4**, p. 137). Sample statements representing this kind of attitude are the following: "Parents cannot help it if their children are naughty." "Quiet children are much nicer than little chatter-boxes."

All scoring of responses followed Shoben's weighting system, a procedure he had found to result in parents of problem children making significantly higher scores, on the average, than parents of nonproblem children. Scores were interpreted as follows: The higher the score, the more intense is the attitude; the lower the score, the less intense is the attitude.

Results

As Table 1 indicates, the psychology students expressed a significant decrease in the intensity of their attitudes, as measured by the total scale. The mean change was 7.63 scale points. This shows that in general the attitudes of the psychology students toward parent-child relationships became more permissive by the end of the course.

The psychology classes also changed significantly in each of the three attitude dimensions which made up the total scale. Table 1 shows that the greatest change was in dominating attitudes, with a mean decrease of 5 scale points. Changes in possessive and ignoring attitudes were smaller, being 1.51 and 1.12 scale points respectively.

These changes were evidently the result of taking child psychology, since an analysis of the sociology classes showed no significant change in attitudes. Table 2 reveals this lack of change between the pretest and posttest scores. A comparison of the initial attitudes of the sociology and psychology students shows that both groups expressed essentially the same intensity of attitudes at the beginning of their respective courses. Only the psychology students, however, changed significantly. This fact supports the conclusion that the psychology students changed as

TABLE 1

ATTITUDES TOWARD PARENT-CHILD RELATIONSHIPS BEFORE AND AFTER COURSE IN CHILD PSYCHOLOGY

(N = 157)

Attitude measured	Before course		After course		Difference in means	t
	Mean score	*SD*	Mean score	*SD*		
Dominating	154.55	11.23	149.55	10.18	5.00	7.46*
Possessive	73.50	6.30	71.99	5.82	1.51	3.75*
Ignoring	52.40	3.87	51.28	3.86	1.12	3.29*
Total scale	280.45	16.30	272.82	15.93	7.63	7.00*

Note.—The higher the score, the more intense is the attitude.

* $p < .01$.

TABLE 2

ATTITUDES TOWARD PARENT-CHILD RELATIONSHIPS OF STUDENTS WHO DID NOT TAKE COURSE IN CHILD PSYCHOLOGY

(N = 155)

Attitude measured	Before course		After course		Difference in means	t
	Mean score	*SD*	Mean score	*SD*		
Dominating	154.36	11.12	154.24	12.29	.12	.17
Possessive	74.44	5.66	74.21	5.04	.23	.56
Ignoring	51.41	4.45	51.58	4.27	.17	.45
Total scale	280.05	16.23	280.21	16.74	.16	.14

Note.—The higher the score, the more intense is the attitude.

TABLE 3

ATTITUDES TOWARD PARENT-CHILD RELATIONSHIPS OF STUDENTS ACHIEVING IN UPPER AND LOWER HALVES OF COURSE IN CHILD PSYCHOLOGY

	Before course		After course		Difference in means	t
	Mean score	*SD*	Mean score	*SD*		
Upper half (N = 79)	275.84	14.56	268.54	14.46	7.30	5.84*
Lower half (N = 78)	285.65	16.11	277.29	15.93	8.36	5.61*

Note.—The higher the score, the more intense is the attitude.

* $p < .01$.

TABLE 4

MEN AND WOMEN STUDENTS' ATTITUDES TOWARD PARENT-CHILD RELATIONSHIPS BEFORE AND AFTER COURSE IN CHILD PSYCHOLOGY

	Before course (total scale)		After course (total scale)		Difference in means	t
	Mean score	*SD*	Mean score	*SD*		
Men (N = 65)	281.05	15.83	272.79	15.73	8.26	5.66*
Women (N = 92)	280.14	16.50	273.02	15.91	7.12	4.98*

Note.—The higher the score, the more intense is the attitude.
* $p < .01$.

a result of taking a course in child psychology, and not simply because of general college experiences.

In order to discover what relationship might exist between scholastic achievement in child psychology and changes in attitudes, a comparison was made between students achieving in the upper half of their class and those achieving in the lower half. The basis for comparison was scores obtained on objective examinations covering course content, including a comprehensive final examination. Results of this analysis are summarized in Table 3. Initially the upper half of the class was more permissive than the lower half. Both groups changed significantly. The amount of change, however, for each group was the same. (The actual difference of 1.06 has a t value of .59.)

Did men and women students differ in their attitude changes? The answer to this question can be found in Table 4. Initially, the attitudes of both groups were approximately the same. Both men and women showed a significant decrease in attitude intensity. The amount of change, however, for each group was the same. (The actual difference of 1.14 has a t value of .59.)

DISCUSSION

At the conclusion of a course in child psychology, students expressed more permissive attitudes toward parent-child relationships than they had held at the beginning of the semester. While they changed significantly in all three kinds of attitudes measured, their dominating attitudes decreased the most. Why did the greatest change occur in that particular area? There are several factors which should be considered in answering this question. All of them, to some extent, probably played a part in effecting change.

First, it may be that for the kind of population which made up the psychology classes, dominating attitudes are more susceptible to change through formal education than are possessive and ignoring attitudes. Most of these students were not very far removed in time and memory from their adolescence. Since problems of parental domination are so outstanding in typical adolescent-parent relationships, it may be that recent sensitization made these students more keenly aware of such problems. Thus they would be ready to react to any course content which depicted the effects overdominating parents have on children.

A second important factor is that the students may simply have examined more course content dealing with dominating attitudes, and thus were able to change more. Of course, this possibility is closely related to the first. They may have encountered more material dealing with parental domination because they were more

ready to react to it, and therefore sought it out more often than information dealing with possessive and ignoring attitudes. (The instructor of the course is satisfied that neither his lectures nor assignments were more heavily weighted with material dealing with dominating attitudes than with other kinds of parent-child relationships.)

A third likely reason for the greatest change being in dominating attitudes lies in the nature of the attitude scale. It will be recalled that the dominating area contained 40 items, as compared with 20 for possessive and 15 for ignoring attitudes. Thus, the dominating area of the scale afforded a wider range of possible responses, and therefore more opportunity to express change. Furthermore, this part of the scale had a higher correlation with the total scale (.86) than did the other two attitude categories (**4**, p. 131). It is also significant to note that when Shoben compared the attitudes of parents of problem children with parents of nonproblem children, he found the greatest difference in the two groups to be in dominating attitudes (**4**, p. 132).

The fact that low achievers changed as much in their attitudes as high achievers confirms something teachers have strongly suspected for a long time: Grades don't tell them all they would like to know about what students obtain from a course! However, had measurements of scholastic progress other than objective examinations been used, the relationship between change in attitude and achievement might have been different.

Evidently the course had the same effect on the attitudes of men and women students, both groups becoming more permissive to the same extent. This is a desirable outcome, since it is a commonplace that both parents play important parts in affecting their children's development.

Considering the great amount of deliberate emphasis on parent-child relationships built into the course, the investigator had anticipated greater changes in attitudes than those discovered. But perhaps it is unrealistic to expect in one semester much more change than this. Then, too, these measured changes may be only the initial stimulation for greater changes which will take place later in the students' development.

In any case, however, studies like this one would probably be more meaningful if they measured more specific attitudes than those represented by Shoben's threefold classification. The investigator is now in the process of doing this by using an inventory recently developed by Schaefer and Bell (**3**). Their instrument is an improvement over Shoben's in several respects, an important one being that its attitude areas, delineated on the basis of factor analysis, are more specific and homogeneous than the areas used by Shoben. It is hoped that others interested in measuring attitudinal outcomes of child psychology will also try out this instrument.

Summary

Will an undergraduate course in child psychology change attitudes toward parent-child relationships? To answer this question, an attitude scale was administered to 157 students before and after the course. Postcourse attitudes were significantly less dominating, possessive, and ignoring. The greatest change was in dominating attitudes. Scholastic achievement in the course was not significantly related to the amount of change. Men and women did not differ significantly in how much they changed. As a control measure, the same scale was administered to 155 students before and after a course in introductory sociology. No change in attitudes occurred.

REFERENCES

1. Birney, R., & McKeachie, W. The teaching of psychology: A survey of

research since 1942. *Psychol. Bull.*, 1955, **52**, 51–68.
2. HURLEY, R., & LAFFEY, J. Influence of a conventional child psychology course upon attitudes toward children. *Papers of the Mich. Acad. Sci., Arts, and Letters*, 1957, (1956 meeting), 299–306.
3. SCHAEFER, S., & BELL, R. Q. Parental attitude research instrument (PARI). Bethesda, Md.: Nat. Inst. of Mental Health, Child Development Section, no date (Mimeo.).
4. SHOBEN, E. J., JR. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monog.*, 1949, **39**, 101–148.

Received October 7, 1957.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 1, 1958

THE DIFFERENTIAL EFFECT OF MANIFEST ANXIETY ON TEST PERFORMANCE[1]

EVAN W. PICKREL

Air Force Personnel and Training Research Center

Though the Taylor-Spence anxiety-drive concept has received considerable attention in the experimental literature, there is a significant lack of information pertaining to the effect of this variable on psychometric performance. The thesis of this study, in accordance with evidence available in the experimental literature, is that manifest anxiety level may be shown to have a differential effect on various phychometric measures; and further, that this differential effect may be shown to vary with the number of alternatives available for decision-making, the amount of habit interference possible within the task. The anxiety-drive hypothesis essentially states that the higher the anxiety score on the manifest anxiety scale (**6**), the stronger the excitatory potential and the greater the response strength. For example, a group scoring high on the scale has been demonstrated (**3**, **4**, **5**) to be superior to a low scoring group in the amount of conditioning exhibited. In more complex learning situations, however, in which there are a number of alternative competing tendencies, it has been shown (**1**) that the effect of increasing a drive would depend upon the initial response hierarchy and the relative habit strength of the correct or goal attaining response in the hierarchy. Differences in drive level may lead to superior performance by either the anxious or the nonanxious group depending on the difficulty of the choice points.

It is therefore assumed that there would be no significant difference in the performance of anxious and nonanxious *Ss* on simple speeded measures which pose problems not having alternative competing tendencies. However, when the test problems offer a minimal number of alternative competing tendencies and thus are parallel to the experimentalist's conditioning problem in terms of possible habit interference, the performance of anxious *Ss* will be superior to that of nonanxious *Ss*. Furthermore, when the measures become quite complex and each problem offers a number of alternative competing tendencies, the performance of nonanxious *Ss* will be superior to that of anxious *Ss*.

METHOD

Measures

Four measures were selected to test the hypotheses. Answer Sheet Marking Test B1710AX was chosen as a simple repetitive task which does not bring forth any alternative competing choices for solution of the problems. The test is used to determine how quickly and accurately *Ss* can locate and mark designated responses on a 15-choice IBM answer sheet. There are two parts, each presenting 75 randomly distributed items.

Army Clerical Speed ACS-2 was selected as a measure in which the test problems offer a minimal number of alternative competing tendencies. The test presents 125 four- to seven-digit numbers paired with either their reversals or a reasonable approximation. The task is to state

[1] This report is based on work done under ARDC Project No. 7719, Task No. 17011, in support of the Research and Development Program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

whether an exact reversal of the stimulus number is presented.

A Code Learning Test was developed to parallel the complex learning situations often used in experimental laboratories. The test utilizes a tape recorded and IBM answer sheets for group administration, and teaches the first 10 letters of International Morse Code by presenting each signal twice, testing at the end of the sequence, and repeating the procedure 10 times. Part scores were obtained for each of the 20-item subtests. All *Ss* reporting previous experience with the code were eliminated from the experiment.

A simple, nonspeeded arithmetic achievement test was selected as a complex measure of overlearning. California Achievement Test II, a subtest 4, section D of parallel forms AA, BB, and CC were used. Each form presents 20 addition problems varying in complexity from the summation of: 2 two-digit to 4 four-digit whole numbers, decimals, fractions and complex numbers when presented in columnal and linear array. Scores were obtained for each parallel form of the test.

Subjects

All *Ss* were newly inducted basic airmen entering Lackland Air Force Base from November 29, 1955 to January 20, 1956. A total of 723 *Ss* were administered a modified form of the Taylor scale which included the 50 anxiety items of the original scale plus the Lie and Manic scales from the Minnesota Multiphasic Personality Inventory. The Lie scale was used to detect false scores, a score of eight or above eliminating an *S* from selection. The anxious and nonanxious groups each consisted of those whose scores fell respectively in the upper and lower 20% of scores for a standardization population of about 1000 basic airmen. The cutoffs were 21 and higher, 10 and lower respectively on the manifest anxiety scale. There were 178 *Ss* in the anxious group and 159 *Ss* in the nonanxious group.

Anxious and nonanxious *Ss* from each flight of approximately 60 men were presented the above measures in the order described.

Results and Discussion

The means and standard deviations for anxious and nonanxious groups on the Answer Sheet Marking and Army Clerical Tests are presented in Table 1. The analysis of variance technique was used to test for significant differences between performances of the two groups on the complex learning and achievement measures. This information is presented in Table 2.

In all cases the hypotheses were substantiated. There was no significant difference in the speed and accuracy with which anxious and nonanxious *Ss* could locate and blacken specified spaces on an IBM answer sheet. The mean performance score for anxious *Ss* was significantly higher than that of nonanxious *Ss* on the Army Clerical Speed Test. On the more complex activities, the Code Learning Test and the arithmetic achievement tests, the mean performance scores of nonanxious *Ss* were significantly higher than those of anxious *Ss*.

Performance on the Army Clerical Speed Test parallels the results obtained with conditioning problems in the experimental laboratories. On the more complex activities, the code learning and arithmetic achievement tests, performance parallels the results obtained with more complex learning situations in the experimental laboratories. Since these results parallel those found with conditioning and complex learning situations, the appropriate placement of the control task on a complexity continuum is open to question. The amount of habit interference possible within this task does not seem as great as within the Army Clerical Speed Test. Yet perform-

TABLE 1
MEAN PERFORMANCE SCORES

Measure	Groups				t
	Anxious (N = 178)		Nonanxious (N = 159)		
	Mn	*SD*	Mn	*SD*	
Answer Sheet Marking	109.51	23.92	112.04	23.24	.98
Army Clerical Speed	70.70	27.75	63.34	26.47	2.48*

* Significant at the .02 level.

TABLE 2
ANALYSIS OF VARIANCE FOR TESTING SIGNIFICANT DIFFERENCES BETWEEN PERFORMANCE OF THE TWO GROUPS

Source	Code Learning Tests		Calif. Achiev. Tests	
	df	Variance	*df*	Variance
Between groups	1	253.09*	1	114.08*
Between trials	9	1396.94*	2	130.99*
Between individuals within groups	335	105.87*	335	40.02*
Groups by trial interaction	9	3.35	2	2.02
Residual	3015	5.29	670	2.96

* Significant at the .01 level.

ance on the Army Clerical Speed Test paralleled the results of conditioning experiments. Either the control task is misclassified on a complexity continuum or the Taylor Scale does not function as a measure of drive in this situation. A study has been initiated to provide the answer to this question.

Previous studies have shown that an extremely high scoring group on a manifest anxiety scale is constantly superior to a low scoring group in the amount of conditioning exhibited but that in more complex learning situations a high drive level could result in an impairment of performance. In this study a control task was also introduced, a situation in which there was no significant difference between the performance of anxious and nonanxious *Ss*. Thus the differential effect of manifest anxiety on performance was shown with psychometric variables to be a function of the number of alternatives for decision-making given to *S*. Since results were obtained which are consistent with those found in experimental literature, the study has demonstrated a parallelism between psychometric measures and some types of variables studied in experimental laboratories.

SUMMARY

Groups of anxious and nonanxious *Ss*, as determined by a scale of manifest anxiety, were administered psychometric tasks which varied in the number of alternatives for decision-making given to *S*. There was no significant difference in the performance of the two groups on a control measure which did not provide *S* with alternatives for decision-making. When the task provided a minimal number of alternatives, mean performance of the anxious group was superior. When the task was complex and a number of alternatives were made available, mean performance of the non-

anxious group was superior. Results were obtained which are consistent with those found in the experimental literature, thus demonstrating a parallelism between psychometric measures and some types of variables studied in experimental laboratories.

REFERENCES

1. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
2. Sarason, S. B., Mandler, G., & Craighill, P. C. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.
3. Spence, K. W., & Taylor, Janet. Anxiety and strength of the USC as determiners of the amount of eyelid conditioning. *J. exp. Psychol.* 1951, **42**, 183–188.
4. Spence, K. W., & Taylor, Janet. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.
5. Taylor, Janet. The relation of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
6. Taylor, Janet. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Received October 11, 1957.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 1, 1958

SOME SOCIOECONOMIC CORRELATES OF ACADEMIC APTITUDE

KHOSSROW MOHANDESSI AND PHILIP J. RUNKEL
Bureau of Educational Research, University of Illinois

It is usually agreed that the socioeconomic status of the family and community in which children and youth grow up affects their learning behavior and school achievement (e.g., **2**, **3**, **5**, **10**). Davis, for example, explains the influence of social class in these words: "By defining the group with which an individual may have intimate clique relationships, our social class system narrows his training environment. His social instigations and goals, his symbolic world and its evaluation are largely selected from the narrow culture of that class with which he can associate freely" (**2**, p. 609).

In the present study, academic aptitude of students in Illinois high schools is compared with some of the socioeconomic characteristics of various communities in which the schools are located. Variables chosen for this purpose were those for which there was ready-to-hand information and which tended to "narrow the culture" of the people in the community. They are the distance from large towns and cities, community population, distance from the nearest active coal mine, the value of farm products sold, the value of land and building per farm, the school size, and whether the school is public or private. Obviously, these are not direct measures of the psychological effects of family and community but rather are rough indices which we presume will be associated with psychological processes which in turn are related to academic aptitude. While these variables are not of the kind to please the theoretician, they have the advantage for the practical worker that the necessary information is readily available.

Although this study was made only of schools within the state of Illinois, there is evidence that the relationships reported here would be typical of those throughout the nation. In a nationwide sampling of high schools, Mollenkopf and Melville (**7**) found that academic aptitude and achievement were related to such variables as region (South or non-South), community size, percentage of fathers who were high school graduates, instructional support per pupil, percentage of support from state aid, and whether the region served by a school had a public library.[1]

Somewhat similar results were obtained by Thorndike (**8**) in his study of community variables as predictors of intelligence and academic achievement. Correlations between IQ and various indices of the education of adult population ranged from .33 to .43 and between IQ and different measures of cost of housing from .30 to .32. A correlation of .28 was obtained between IQ and proportion of native-born whites, one of −.26 between IQ and rate of female employment, and one of .28 between IQ and frequency of professional workers in community.

In the present study, the academic aptitude of the students was measured by the portion of the Differential Aptitude Tests (**1**) which is used as a part of the Illinois Statewide High School Testing Program. The score used is the Total score from the Abstract Reasoning and Verbal Reasoning Tests. This score will be referred to as the "D.A.T." in this paper. We will also talk of the Illinois Statewide Testing Program as the "Program."

From the 1955–56 Program we have figures available showing the mean D.A.T. for each school computed from the per-

[1] This prepublication memorandum is cited with the kind permission of the authors.

centile scores of the pupils in that school. Using this information, we chose as our "dependent" variable an extreme dichotomy; namely, whether the school mean lies in the top quarter of all schools or in the bottom quarter. The second and third quarters are ignored.[2] It should be remembered that not all high schools in the state appear in the Program and correspondingly in this analysis. The schools included are those participating in the Statewide Testing Program and having school means on the D.A.T. falling either in the highest or the lowest quarters.

Hypothesis

The purpose of this study was to check some rough indices of the socioeconomic characters of different communities in the state of Illinois (or of the schools which are located in those communities) against the mean academic aptitude of the students sent to schools in those communities. Our general hypothesis is that *some of the factors which determine aptitudes in schools in various communities are to be found among the socioeconomic characteristics of those communities.* We would suppose that, while so e of this relationship is due to the efficacy with which certain types of social situations foster the growth of academic ability, another part of the relationship is due to the self-selection of the kinds of families which move into the various kinds of communities. Our prediction in each case below stems from the expectation that the more a community is able to support its schools, the greater are the chances that the students will show the higher ranges of aptitude.

[2] Since cases falling in the second and third quarters are not examined, conclusions cannot be drawn about these quarters. We can only draw conclusions about whether a variable (such as distance from larger towns) is associated with the school's falling in the *first* or *fourth* quarter.

TABLE 1
Mean D.A.T. Versus Distance from Nearest Town of 25,000 or Larger (N = 167)

Distance of schools from towns of 25,000 or larger	Quarter of mean D.A.T. of school	
	First	Fourth
Less than 10 miles	47	12
Over 10 to 25 miles	35	20
More than 25 miles	12	41

Note.—$p < .001$ by chi square for 2 *df*.

Findings

A description of the socioeconomic measures used in the study and the results obtained in each case follow.

Distance from Town of 25,000 or Larger

First examined was the distance of each school from the nearest town of 25,000 or more in population. These distances were scaled from an ordinary current highway map. Communities farther from the larger towns, we felt, would be more rural in character. Table 1 shows the distribution of schools in the first and last quarters of the mean D.A.T. scores according to their distance from the nearest town of 25,000 or more. The mean distance from such towns is 12 miles for the schools in the first (highest) quarter of mean D.A.T. scores and 29 miles for the schools in the fourth (lowest) quarter. The chi-square test applied to the data of Table 1 gives a value significant at the .001 level and supports the prediction made.[3]

It should be noted, in regard to Table 1 and all other tables, that we chose the categories of the independent variable so

[3] We report in each case the p value which is the smallest the table lists for the obtained value of chi square. We report these values so that the reader may choose for himself the level of probability which he wishes to consider statistically significant.

as to keep the frequencies in each category as equal as might be while at the same time avoiding expected frequencies too small to satisfy the requirements of the chi-square test.

Community Size

A second variable used was the population of the communities in which the schools with higher and lower D.A.T. means were located. The assumption was that larger centers of population would be able to spend more money on their schools than the smaller ones. The relation was examined twice: once including schools in the Chicago metropolitan area and a second time excluding them. Population figures were taken from the highway map. Towns not listed on the map were assumed to have populations under 1,000. As shown by Tables 2(a) and 2(b), the expected relation is significant at the .001 level whether or not Chicago schools are included.

Distance from Nearest Active Coal Mine

The next prediction was that proximity to coal mines would tend to lower the mean aptitude to be found in the school. Table 3 shows the distribution of schools in the first and fourth D.A.T. quarters according to their distances from the nearest active coal mine. This variable was chosen because a great number of people, especially in the southern part of the state, make their living through working in the coal mines. Some of the farmers work in them to earn "extra" money. Thus, existence of coal mines becomes an important factor in shaping the economy of an area and in deciding the kind of population inhabiting certain localities, particularly in the southern half of the state. The appropriate information was taken from a map of mineral industries (**4**). The relation predicted is significant at the .001 level.

TABLE 2
MEAN D.A.T. VERSUS SIZE OF COMMUNITY

Population of the community where schools are located	Quarter of mean D.A.T. of school	
	First	Fourth
(a) Including Chicago Schools[a] (N = 167)		
Under 1,000	14	33
1,001–5,000	25	28
5,001–25,000	27	10
Over 25,000	28	2
(b) Excluding Chicago Schools[b] (N = 144)		
Under 1,000	14	33
1,001–5,000	25	28
Over 5,000	33	11

[a] $p < .001$ by chi square for 3 *df*.
[b] $p < .001$ by chi square for 2 *df*.

TABLE 3
MEAN D.A.T. VERSUS DISTANCE FROM NEAREST ACTIVE COAL MINE
(N = 167)

Distance from nearest active coal mine	Quarter of mean D.A.T. of school	
	First	Fourth
Less than 20 miles	16	31
Over 20–35 miles	18	20
Over 35–50 miles	33	11
Over 50 miles	27	11

Note.— $p < .001$ by chi square for 3 *df*.

Value of the Farm Products Sold in the County

The value of the farm products sold in different counties of the state was used as an index of the income of the residents of a given locality. The analysis was made separately for counties containing towns of 25,000 or over and for counties not containing such towns in order to find out whether there was any difference between the two kinds of counties in this respect. It was thought that where value of farm

products sold was higher, urban and rural areas would both prosper and property values would be high in general. It was expected that schools with the higher D.A.T. means would be located in communities which showed the greater values of farm products sold during the year. Values were taken from the 1954 U. S. Census of Agriculture (**9**). Tables 4(a) and 4(b) present the relevant data. The expected relation was significant for both categories, i.e., counties with and without larger towns, beyond the .05 level.

TABLE 4
MEAN D.A.T. VERSUS VALUE OF FARM PRODUCTS SOLD

Value of farm products sold in millions of dollars	Quarter of mean D.A.T. of school	
	First	Fourth
(a) Counties Containing Towns of 25,000 or Larger[a] (N = 68)		
0–14.99	5	7
15.0–24.99	30	4
25.0 and above	17	5
(b) Counties Not Containing Towns of 25,000 or Larger[b] (N = 99)		
0–14.99	15	29
15.0–24.99	12	20
25.0 and over	15	8

[a] $p < .01$ by chi square for 2 *df*.
[b] $p < .05$ by chi square for 2 *df*.

TABLE 5
MEAN D.A.T. VERSUS AVERAGE VALUE OF LAND AND BUILDINGS PER FARM

Value of land and buildings per farm in thousands of dollars	Quarter of mean D.A.T. of school	
	First	Fourth
(a) Counties Containing Towns of 25,000 or Larger (N = 68)		
Less than 45.0	31	8
More than 45.0	21	8
(b) Counties Not Containing Towns of 25,000 or Larger[b] (N = 99)		
Less than 45.0	17	32
More than 45.0	25	25

[a] $p > .10$.
[b] $p > .10$.

Value of Land and Building Per Farm

As a second index of income and therefore of the ability of rural people to give financial support to their schools, the average value of land and buildings per farm was used. This likewise was computed separately for counties containing towns of 25,000 or over and counties not containing such towns. The information was taken from the 1954 U. S. Census of Agriculture (**9**). Tables 5(a) and 5(b) contain the data relevant to this question. The partitioning of chi square (**6**) was used to test the significance of the relations among the variables involved; i.e., quarter of D.A.T., average value of land and buildings per farm, and whether or not the county contained any town of 25,000 and over. The only significant relation—at the .001 level—was that between quarter and location in counties with towns of 25,000 or over. This latter relation is of little interest here since it is only another form of the relation between mean D.A.T. and the distance from larger towns. In brief, our expectation was *not* borne out that the mean D.A.T. of the school would be related to the value of farm land and buildings in the county. We have no ready explanation of the reason that this variable should fail to show a significant relation while our other variables, just as imprecise, do so.

Size of School

Another variable used was the size of the school itself. This information is avail-

TABLE 6
MEAN D.A.T. VERSUS SCHOOL SIZE*
(N = 138)

Number of pupils in school	Quarter of mean D.A.T. in school	
	First	Fourth
1–99	5	23
100–299	28	36
300–499	13	8
500–999	11	4
More than 1,000	10	0

Note.—The information concerning this variable was lacking for 29 schools in the Program.

* $p < .001$ for chi square for 4 *df.* (computed with the Yates correction for continuity)

TABLE 7
MEAN D.A.T. VERSUS PUBLIC-PRIVATE VARIABLE
(N = 167)

Public -private variable	Quarter of mean D.A.T. of school	
	First	Fourth
Public schools	66	71
Private schools	28	2

$p < .001$ by chi square for 1 *df.*

able from the files of the Illinois Statewide High School Testing Program. The prediction was that schools in the first quarter would be significantly larger in size than the schools in the fourth quarter. Table 6 contains the data relevant to this point. The prediction is supported at the .001 level.

The Public-Private Variable

This measure was used as another variable which might correlate with mean aptitude in the school. It was predicted that the ratio of private schools to public schools would be significantly larger in the first quarter of D.A.T. scores than in the fourth quarter. Here one would suppose the effect to be due largely to the selectiveness of the private school. Table 7 presents the relevant data. The frequencies in Table 7 seem at first glance very "lopsided." The reason for this, of course, is that so many more public than private schools participate in the Illinois Statewide Testing Program. Nevertheless, the chi-square value for the relation in Table 7 is significant at the .001 level.

SUMMARY AND CONCLUSION

The purpose of the present study was to investigate some socioeconomic correlates of academic aptitude. The study was intended to be exploratory rather than precise. Academic aptitude was examined by using schools whose mean test scores were in the upper or lower quarters among schools participating in the Statewide Testing Program of the University of Illinois. Findings of the study were that the mean academic aptitude of the school was strongly related to (*a*) the distance from larger towns, (*b*) the community population, (*c*) the distance from active coal mines, (*d*) the value of farm products sold in the county where the school is located, (*e*) the size of the school, and (*f*) whether the school is public or private.

REFERENCES

1. BENNETT, G. K., SEASHORE, H. G., & WESMAN, A. G. *Differential aptitude tests,* Form A. New York: The Psychological Corp., 1947.
2. DAVIS, A. Child training and social class. In R. G. Barker, J. S. Kounin, & H. F. Wright (Eds.) *Child behavior and development.* New York: McGraw-Hill, 1943. Pp. 607–619.
3. DAVIS, A. American status systems and the socialization of the child. In C. Kluckhohn, & H. A. Murray, (Eds.) *Personality in nature, society, and culture.* New York: Knopf, 1949. Pp. 459–468.
4. Educational Extension Division, State Geological Survey, Department of Registration and Education, State of Illinois. *Mineral industries of Illinois.* (Map) Springfield: Author, 1955.
5. EMPEY, L. T. Social class and occupa-

tional aspiration: A comparison of absolute and relative measurement. *Amer sociol. Rev.*, 1956, **21**, 703–709.

6. LANCASTER, H. O. Complex contingency tables treated by the partition of chi-square. *J. Royal stat. Society*, Series B, 1951, **13**, 242–249.
7. MOLLENKOPF, W. G., & MELVILLE, S. D. A study of secondary school characteristics as related to test scores. *Research Bulletin No. 56–6*. Princeton: Educational Testing Service, 1956.
8. THORNDIKE, R. L. Community variables as predictors of intelligence and academic achievement. *J. educ. Psychol.*, 1951, **42**, 321–338.
9. U. S. Bureau of Census. *U. S. census of agriculture:* Vol. 1, Counties and state economic areas, Part 5, 1954. Washington: U. S. Government Printing Office, 1956.
10. WARNER, W. L., HAVINGHURST, R. J., & LOEB, M. B. The social role of the teacher. In T. M. Newcomb, & E. L. Hartley, (Eds.) *Readings in social psychology*. New York: Henry Holt, 1947. Pp. 479–480.

Received October 24, 1957.

Publication Manual

of the American Psychological Association

1957 Revision

A revision of the 1952 Manual, detailed instructions are given for the preparation of scientific articles. Organization and presentation of tabular material, figures and graphs, and reference lists are included. All scientists who are writing for publication will find the Publication Manual an indispensable guide.

Price, $1.00

Discounts for quantity orders over fifty copies

Order from

AMERICAN PSYCHOLOGICAL ASSOCIATION

Publications Office
1333 Sixteenth Street, N. W.
Washington 6, D. C.

Contemporary Psychology

A JOURNAL OF REVIEWS

Edwin G. Boring, Editor

HAVE YOU READ THESE REVIEWS?

The Clinical Psychologist, by W. A. Hunt

Reviewer: G. A. Kelly

Basic Psychology, by Leonard Carmichael

Reviewer: C. T. Morgan

Sense and Nonsense in Psychology, by H. J. Eysenck

Reviewer: Ruth S. Tolman

Theories of Personality, by C. S. Hall and Gardner Lindzey

Reviewer: H. L. Ansbacher

School Problems in Human Relations, by Lloyd Cook and Elaine Cook

Reviewer: S. M. Corey

Subscription, $8.00

Foreign, $8.50

Single Copy $1.00

AMERICAN PSYCHOLOGICAL ASSOCIATION

Publications Office

1333 Sixteenth Street, N.W.

Washington 6, D. C.

Psychology and Mental Health

A report of the Institute on Education and Training for Psychological Contributions to Mental Health, held at Stanford University in August, 1955.

Edited by CHARLES R. STROTHER

Topics discussed at the Institute:

Training Needs of Psychologists in Community Mental Health

Specialization in Training

Practicum Training

Training for Therapy

Training for Research in the Mental Health Field

Problems of Departmental Organization

Price: $1.75

Order from:

American Psychological Association, Inc.
1333 Sixteenth Street, N. W.
Washington 6, D. C.

IMPROVED
SERVICES
NEW TESTS
NEW INSIGHTS
FROM

The Picture World Test
CHARLOTTE BÜHLER and MOSES P. MANSON
The Grayson Perceptualization Test
HARRY M. GRAYSON
The Sherman Mental Impairment Test
MURRAY H. SHERMAN
The Forer Structured Sentence Completion Test (Forms for Adults and Adolescents)
BERTRAM R. FORER
The Forer Vocational Survey
BERTRAM R. FORER
The Cassel Group Level of Aspiration Test: Revised
RUSSEL N. CASSEL
The M A C C Behavioral Adjustment Scale
ROBERT B. ELLSWORTH
The H T P Children's Revision: Post-Drawing Interrogation Form
ISAAC JOLLES
The H T P Bibliography
V. J. BIELIAUSKAS
The Thalmann Arithmetic Knowledge and Information Test
H. R. THALMANN
Fibre Built Doll Play House
Plastic Doll Play Family

Hundreds of
new current books
Teaching aids
Apparatus

New 1958-1959 Catalogues Sent upon Request

WESTERN
PSYCHOLOGICAL
SERVICES

10655 SANTA
MONICA BOULEVARD
LOS ANGELES 25,
CALIFORNIA